DOCUMENT RESUME

ED 209 308 TM 810 828

AUTHOR Davis, Ann E.; And Others;
TITLE Interpreting Title I Evaluation Results.
INSTITUTION Northwest Regional Educational Lab., Portland, Oreg.
PUB DATE [81]
NOTE 95p.; Distributed by the Technical Assistance Centers.

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Comparative Analysis; Elementary Education; *Information Utilization; *Program Evaluation; Reading Teachers; *Scores; *Workshops
IDENTIFIERS *Elementary Secondary Education Act Title I; *Normal Curve Equivalent Scores

ABSTRACT

This is a workshop simulation and interpretation guide designed for Title I teachers and district personnel. The participants should have some experience with the norm-referenced evaluation model. They learn to interpret normal curve equivalents (NCE) and NCE gains. Participants are led through an interpretive hierarchy from simple, descriptive interpretations of NCE gains to comparisons of results to other standards. The guide gives a brief introduction to exploratory/diagnostic and comprehensive interpretations. The simulation provides a complete evaluation report for a Title I project. A presenter's guide provides detailed information for the workshop leader. Specific objectives for participants include description of meaning of NCE gain, conversion to a percentile change, construction of confidence bands around gains and use of them to guide interpretations. Other areas are identification of comparisons for NCE gains, determination of gain difference from a standard determination of whether a project needs more thorough investigation and where to focus it, and identification of areas to gather information. (DWH)

ESEA Title I Evaluation

Technical Assistance Centers

INTERPRETING TITLE I EVALUATION RESULTS

INTERPRETING TITLE I EVALUATION RESULTS

DEVELOPED BY:

Ann E. Davis
Dennis Deck
Randy Demaline
Suzanne Hiscox

Distributed by: TAC Regions 8, 9 and 10
Northwest Regional Educational Laboratory
710 S.W. Second Avenue
Portland, Oregon 97204
(503) 248-6971

CONTENTS

INTERPRETING TITLE I EVALUATION RESULTS

INTERPRETING TITLE I EVALUATION RESULTS

Section I: Workshop Guide

Synopsis

INTERPRETING TITLE I EVALUATION RESULTS

Rationale

The purpose is to provide participants with a better understanding of the results from a norm-referenced evaluation. These materials help define what interpretations can be made from the end-of-year evaluation report.

The two components of the workshop, the simulation and the interpretation guide are based on an interpretative hierarchy:

1. Descriptive--simple interpretations dealing with what an NCE gain signifies.

2. Comparative--interpretations based on comparing the results to other standards

3. Exploratory/Diagnostic--interpretations that involve more thorough descriptions of program components and achievement impact

4. Comprehensive--interpretations that expand beyond the achievement impact such as student attitude effects and the effects on others besides students.

The activities in this workshop concentrate on the first two levels with a brief introduction into the third level and a mention of the fourth.

Assumptions About the Participants

The intended audience is the Title I teacher. However, District Title I personnel and parents have attended the workshop and found it helpful. Familiarity with the norm-referenced evaluation model is the most important entry consideration. The audience should have some experience with or knowledge of the model and the reporting forms used with the evaluation. If they do not, there is a tendency for the participants to get involved more with questions of model implementation and form completion than with interpretation. Therefore, a participant should have at least one of the following experiences:

- Completion of a Model A orientaton workshop
- Implementation of the evaluation model and completion of report forms
- Attendance at an end-of-year data analysis workshop

Participant Objectives

After this workshop, participants should be able to:

- Describe what an NCE gain means
- Convert an NCE gain to a percentile change
- Construct a confidence band around an NCE gain
- Use confidence bands to guide their interpretations
- Identify some comparisons for an NCE gain
- Determine whether an NCE gain is different from a standard
- Determine whether a project needs a more thorough investigation
- Determine where to focus an investigation of a project
- Identify some areas to gather information

Instructional Activities (1 1/2 hours):

• Introduction to Workshop Components	(5 minutes)
• Introduction to Simulation	(5 minutes)
• Discussion of Descriptive Interpretations	(30 minutes)
• Discussion of Comparative Interpretations	(30 minutes)
• Discussion of Exploratory/Diagnostic Interpretations	(15 minutes)

Materials Needed

- Overhead projector and screen
- Workshop transparencies and blank transparencies
- Workshop materials to distribute to every person

Description

The core of the workshop is the Interpretation Guide that the participants work through with the aid of the simulation activities based on the end-of-year report from a hypothetical school district. It is most effective if participants have recent results from their own evaluations to which they can apply the strategies of the workshop.

Component Description

INTERPRETING TITLE I EVALUATION RESULTS

1. Interpretation Guide for Title I Evaluation Results

This guide is designed to take the participant from simple explanations of what an NCE gain means to more complicated concerns such as where to look for the strengths and weaknesses in the program. To do this, the guide is divided into three major sections.

Section One deals with interpretations of NCEs and NCE gains. The participant will find explanations as well as alternative ways of describing gains. Also, problems of interpreting gains for small class sizes will be discussed.

In Section Two the participant will start with these initial interpretations and add some information by comparing these results to different standards. These standards include last year's results, other projects' results, state averages or objectives set by the program.

Finally, in Section Three the participant will find a discussion of whether further investigation is needed based on substantiating the results and then where to begin looking for more information about a program.

2. Interpreting Title I Evaluation Results: Simulation

A simulation activity designed to take the participants through the sections of the Interpretation Guide and to provide exercise in doing some of the calculations mentioned in the guide.

The major sections of the simulation are:

a. Title I Reading Project Description, Wilson Elementary School, This Year's Achievement Impact Report

b. This Year's Achievement Impact Report for Cherry Hill Elementary School
Last Year's Achievement Impact Report for Wilson Elementary School
Achievement Impact Information from This Year's State Title I Report
Program Objectives for Centennial School District's Title I Reading Project.

c. Graphs for showing Wilson's results

d. Title I Director's Annual Project Summary
Title I Evaluator's Summary

Presenter's Guide for the

Interpreting Title I Evaluation Results Workshop

Materials Participants Should Have at Outset of Workshop

1. Interpretation Guide for Title I Evaluation Results
2. Title I Reading Project Description for Wilson Elementary School.
3. This year's Achievement Information Report for Wilson Elementary School.

Introduction to the Workshop

Introduce Presenters.

Tell participants there were two questions that we've seen in the field that initiated the formation of this workshop: "I wonder what this end-of-year report information means to me?" and "Last year we got a 7 NCE gain and this year we only got 6; is something wrong with our program?"

This workshop will try to help people address these questions. First, we want to provide participants with more information on what the NCE gains mean to them and how they can be used to provide information about a program. Secondly, we would like to help them be able to explain the results to others. This is not only useful to the Title I teacher but that can also be used in explaining results to other audiences like parents, administrators, or school boards. (This second point is important to emphasize because it can provide a participant an excuse to ask a question; it is much easier to ask "How do I explain this to other teachers?" than it is to say "I don't understand it.")

Materials	Strategies
	Explain to the participants that the piece from the workshop that we expect them to use after they return to their schools is the Interpretation Guide. Rather than simply walking through the guide, we are going to use a simulation to explain the different sections of the guide.
Handout 15.207	The brief description of the Wilson Elementary Reading project is to give them a feel that the Wilson project is not unusual. It is not essential that the description be read in detail and the aspects of the Wilson project be memorized.
Transparency No. 1	The first point we should make before we start interpreting the Wilson project is that we are assuming the information from the Wilson project is valid. Page 2 of the Interpretation Guide lists some points which we will assume that Wilson has done.

Materials	Strategies
	(Note to the presenter. One of the problems with the Wilson example is that the test does not match the objectives of the project. Violation of this initial assumption will be discussed later in the simulation.)

Descriptive Interpretations--Simulation

Handout I-H-I
Transparency
Nos. 2, 3

Have participants key in on the bottom right hand corner of Wilson Elementary Achievement Information form for June 15, 1980, where NCE gains are figured. Give them a few seconds to digest the info mation. If there are participants who are unsure about what an NCE is, refer them to pages 3 and 4 of the Interpretation Guide. Some suggested questions and responses are:

- Did the Wilson project have an impact?
  On these students, yes.
- Was the impact positive or negative?
  Positive.
- Should the Wilson teacher be satisfied with these results?
  Some teachers will hedge, but invariably someone will say "Anything above zero is good." Satisfaction is another realm. Teachers will compare it to their experiences, but no one can really declare that it is satisfactory.

Transparency
Nos. 4,5,6

- How would you describe the impact of this project? (Page 5 of the Interpretation Guide reviews the assumptions of the norm-referenced evaluation model and how impact is defined.) Page 7 discusses how to interpret an NCE gain in terms of a percentile change. NCEs have received so much Title I press that we have inadvertently implied that percentiles are just as bad as grade equivalents. For presenting the effect of a Title I project to teachers, administrators, or boards, it might be easier to use percentiles. Put the Wilson example in terms of percentile growth. Caution participants against comparing absolute percentile gains--that's the reason NCEs were devised.

Materials

Strategies

- Was the Title I project effective for all students?
  You can't tell from these numbers. NCE gains here refer to the group average. Some students gained more than this average; some, less; some may even have lost. The NCE gain reflects the average. Another thing you might want to look at is the range of NCE gains for the group of students. This gives you an idea of the extremes of the gains or losses. You might even want to make a histogram showing the gains or calculate the standard deviation of the gains. All of these provide more information to interpret the effectiveness of a project.

- What does the NCE gain tell you about the effectiveness of the project for students like this?
  Refer participants to page 9 of the Interpretation Guide. This starts the introduction to confidence bands around the NCE gain. These bands come from two sources of error:

  1) Measurement Error. To introduce the participants to the idea of confidence bands refer to the score bands with which most teachers are at least faintly familiar from a teacher training measurement course.

  2) Sampling Error. Most teachers have experienced the case where in a small class, almost all students do well except for one or two who do so poorly they pull down the class average to where it doesn't truly reflect the project's effect. This kind of error is most pronounced in small classes.

  Tell the participants that there are two questions that are similar but different enough to make the answers different. The first question is "Was the Title I program effective for this particular group of students?" The answer to this is obvious from the NCE gain on the report. The second question is "Is this Title I program effective for other students like these? If I didn't change anything for next year, what kind of gains could I expect?" The answer to this question involves the use of confidence bands.

Materials	Strategies
Transparency Nos. 7, 8	Figure 3 on page 10 of the Interpretation Guide shows the relation between class size and error. (Transparency No. 7) This figure is translated into a more usable form in Table 1 on page 12. (Transparency No. 8)

To show participants how to use this table (it sometimes helps to call this table "Give-or-Take" Table), use an example from the Wilson report: the fourth grade had a gain of 1.3 NCEs based on 28 students which the table says will have an error of 3.0 NCEs. Therefore, the teacher at Wilson could say that based on the results of this year's evaluation, the effectiveness of the Wilson program for fourth grade students is 1.3 NCEs "give or take" 3.0 NCEs. The first thing that participants will note is that this means that the NCE gain could be anywhere between -1.7 NCEs to 4.3 NCEs. In turn, this means that next year if nothing is changed, there is a chance that the NCE gain could be negative. Ideally, the confidence band should not include zero. Additional note: This is also a good place to reinforce the distinction between "N" and "Membership" on the reporting forms. Sometimes participants want to use "Membership" instead of "N" when using the give-or-take table.

Handout I-H-2

Participants should graph the confidence bands so they get a different picture. Hand out blank graphs for them to work on. Walk them through an example on a transparency.

Transparency Nos. 9, 10 I-H-3, I-H-4

In districts that have small classes, the participants will notice that the error in an NCE gain is very high and they will have to show substantial NCE gains to avoid having zero in the confidence band. Page 15 of the Interpretation Guide gives some suggestions for small classes.

- Combine NCE gains across buildings to look at district NCE gains,
- Combine NCE gains from adjacent grades if program is essentially the same, or
- Combine NCE gains across years.

Materials	Strategies

Comparative Interpretations--Simulation

To help Wilson decide whether the NCE gains should be investigated it may help to look at some comparative information. Handout materials that contain comparative information:

Cherry Hill Elementary's report
Wilson Elementary's report from last year
State average NCE gains
District objectives

Handout 15.272, 15.273, 15.274
Transparency No. 12

- How do Wilson's results compare to Cherry Hill's?

Transparency No. 1
Handout I-H-5

They are higher. Warn participants that when comparing gains, they also have to remember confidence bands. Hand out graphs that have both Wilson and Cherry Hill grade four on them and graph confidence bands based on the give-or-take table. (Transparency No. 8) If the two bands do not overlap, then participants can be more sure that there really is a difference. Use an example from the Wilson data: The actual mean gain for third graders in a project like this could be as low as -1.0 while in the fourth grade it could be as high as 4.4. Therefore, we're not sure whether third grade gains are actually higher or lower than fourth grade gains.

Transparency No. 14
Handout I-H-6

Mention that comparisons should take into consideration the possible differences in student types, administrative, community, and parent support, or other program factors as well as looking for comparability as to the tests used and the evaluation cycles.

- How do Wilson's results compare to last year?

Transparency Nos. 15, 16
Handout I-H-7

Fairly consistent. We could rule out any circumstances peculiar to this school year.

- How do results compare to the State average gains?

Transparency No. 17

They're lower. The same idea of confidence bands holds. The State's gains are based on such a large population, however, that the band will be much smaller. You can graph this also for comparison purposes.

Materials	Strategies
	o How do Wilson's gains compare to the district objectives?
Transparency No. 18	They are lower, but note that Centennial School District has set fairly high goals that could be unrealistic. Previous NCE gains were probably not referenced when stating this objective.

Exploratory/Diagnostic Interpretations--Simulation

The last section of the Interpretation Guide suggests some areas that may be investigated in a project to identify the strengths and weaknesses. This section is only touched on briefly and can be more thoroughly covered in workshops or needs assessment or implementation evaluation.

Transparency No. 19
Handout 15.279, 15.280

As a lead in to the discussion, finish the simulation by handing out the final two pieces: The Director's Summary and the Evaluator's Summary. Ask the participants to work in small groups to identify what they feel may be causes of the results from the evaluation report. Three have been built into the simulation:

1) The objectives of the Cherry Hill project emphasize comprehension more than Wilson's. The test used for evaluation is the reading comprehensive subtest. (Note that an assumption that we made earlier was that the test matched the curriculum. This is a point where the presenter may want to emphasize that tests should be re-examined for content validity on a regular basis.)

2) The Title I teacher spends more time on administrative tasks than with actual teaching.

3) Most of the students already had achieved the objectives in the Wilson program when they entered.

After the groups have worked for about 15-20 minutes, ask them if they've discovered some possible problems. Discuss these in the large group.

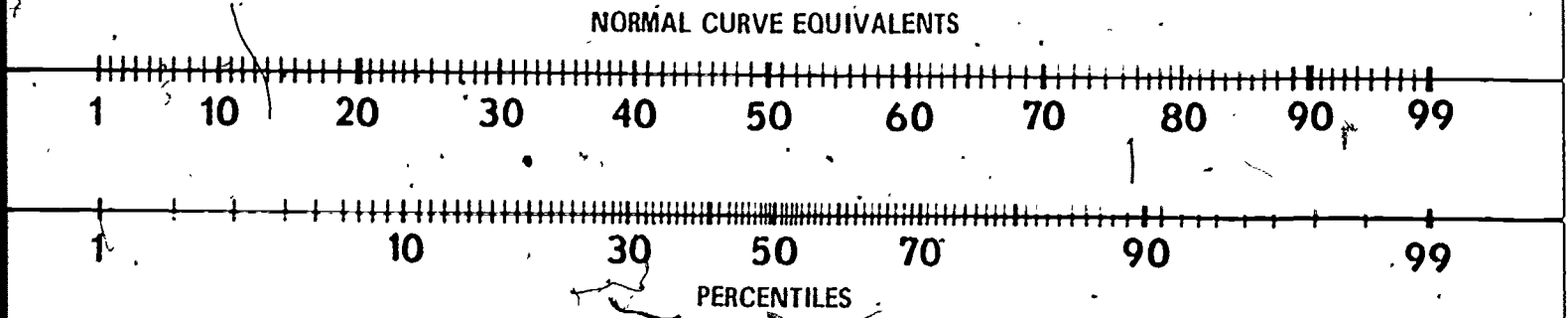
Finish the workshop by discussing what comprehensive evaluation involves. See page 33 of the guide for a brief discussion. Explain that this workshop is not designed to go into this area, but the participants are not to interpret this as meaning that it's not important.

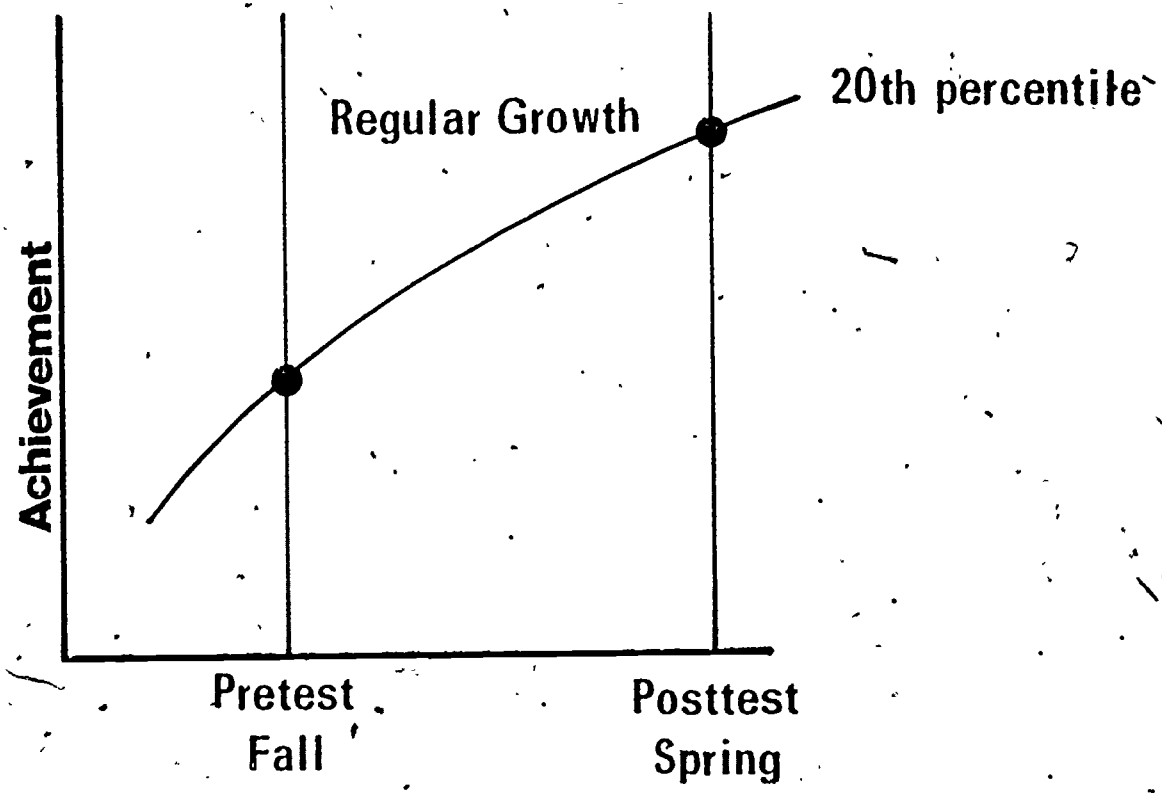
INTERPRETING TITLE I EVALUATION RESULTS.

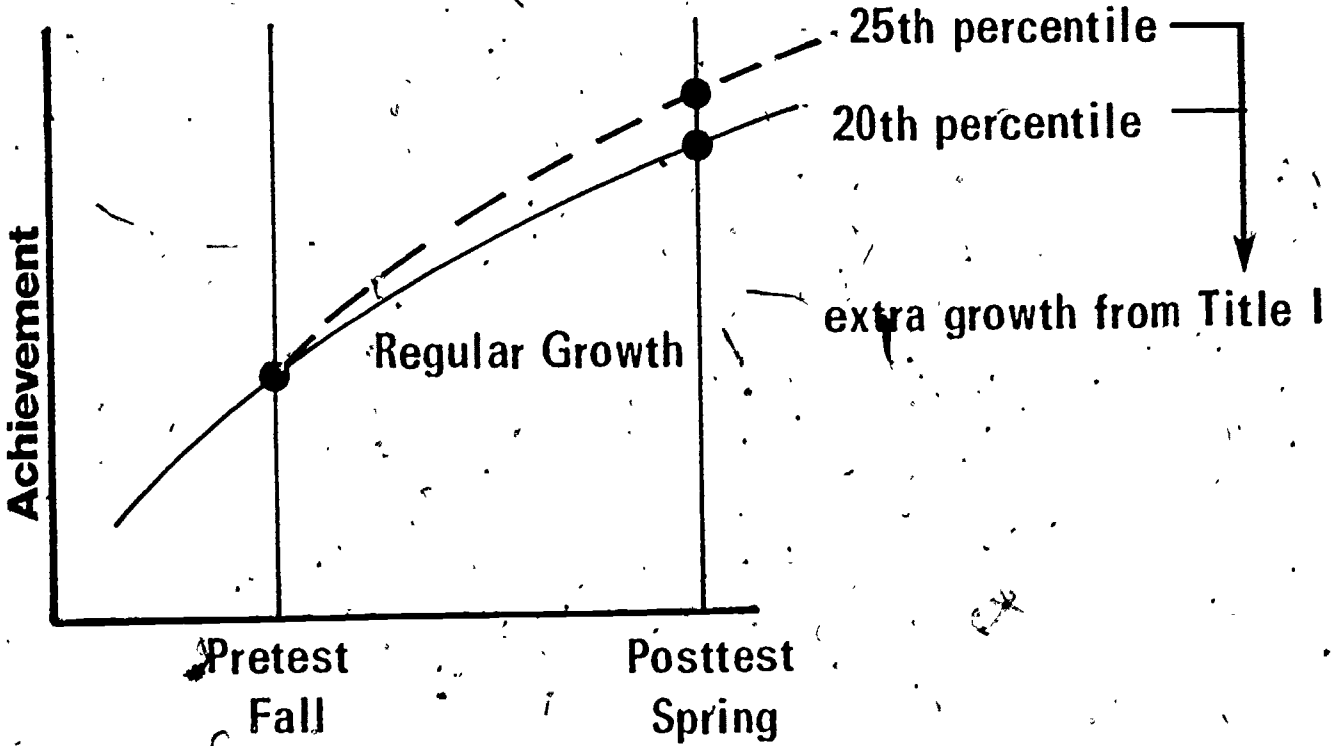
Section II: Overhead Transparencies

- Used Sensitive Test
- Followed Publisher's Instructions
- Controlled Testing Conditions
- Followed Evaluation Model Guidelines
- Met Technical Standards

	NORMAL CURVE EQUIVALENTS (NCEs)		
N	Pretest	Posttest	Gain
27	29.9	32.0	2.1
28	30.8	32.1	1.3
22	27.1	28.8	1.7

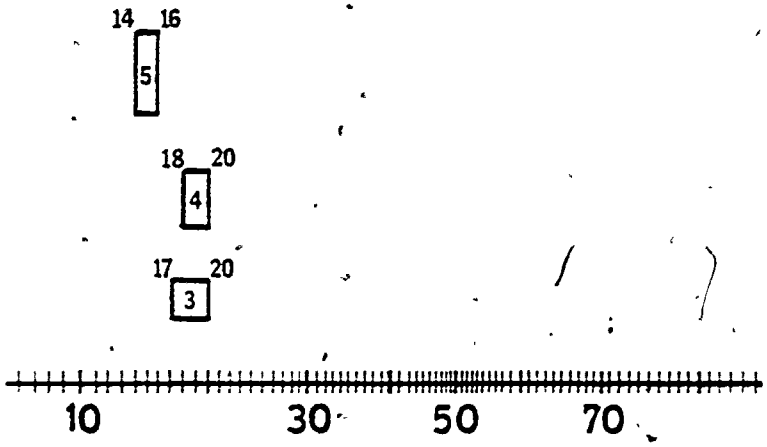

NORMAL CURVE EQUIVALENTS
1
10
20
30
40
50
60
70
80
90
99
1
10
30
50
70
90
99
PERCENTILES

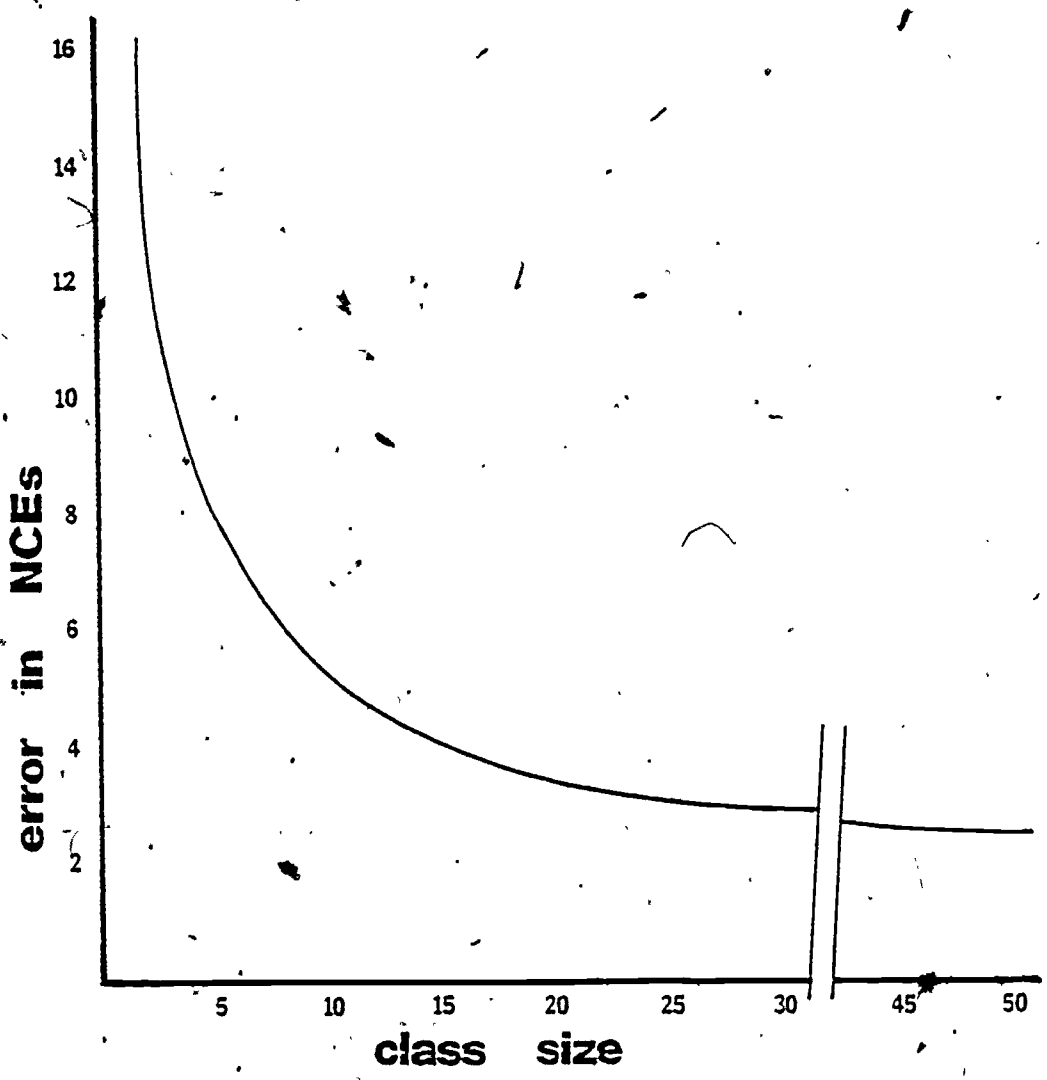

Regular Growth
20th percentile
Achievement
Pretest
Fall
Posttest
Spring


25th percentile
20th percentile
extra growth from Title I
Regular Growth
Achievement
Pretest
Fall
Posttest
Spring

Percentile to Normal Curve Equivalent Conversion Table

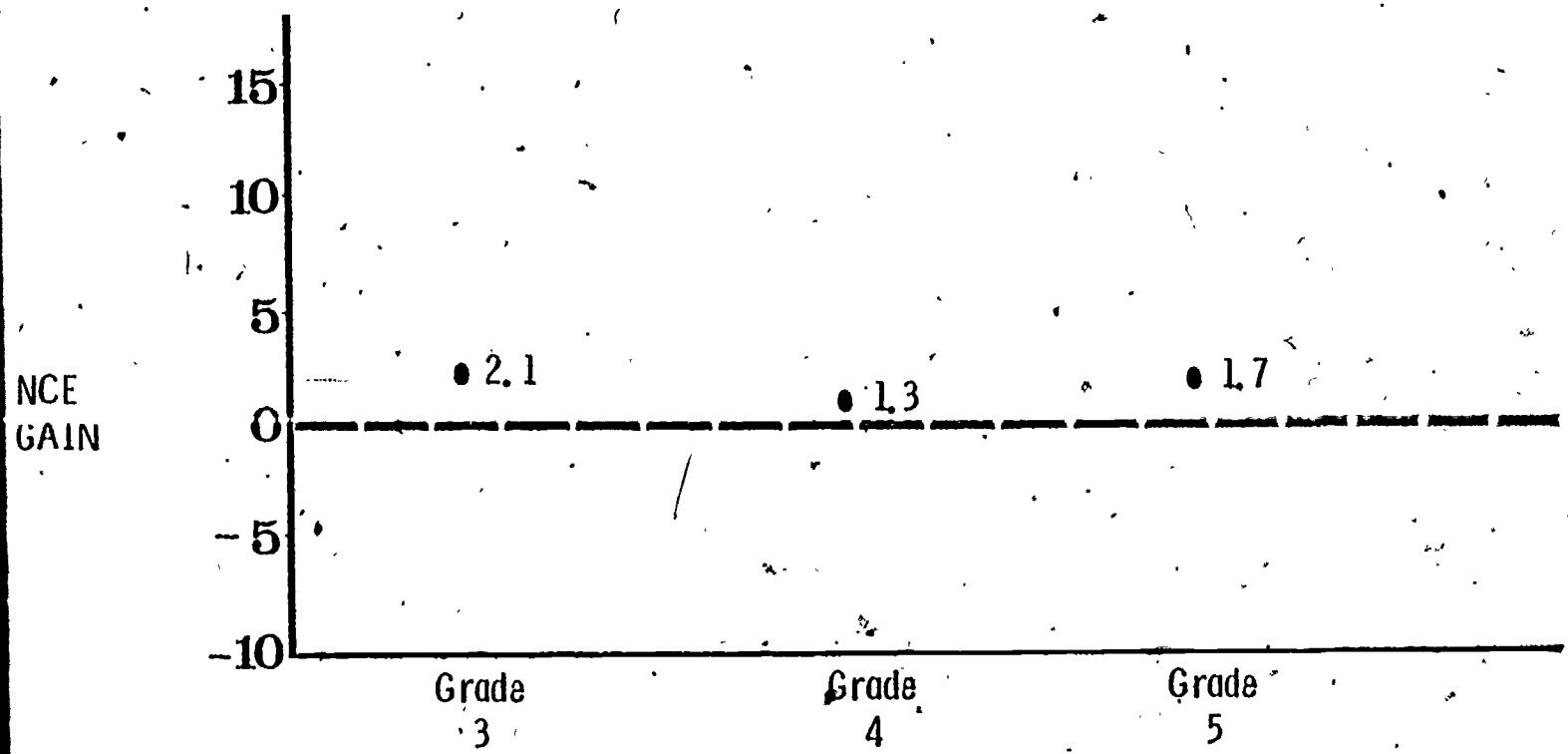
%	NCE	%	NCE	%	NCE
1	1.0	41	45.2	71	61.7
2	6.7	42	45.8	72	62.3
3	10.4	43	46.3	73	62.9
4	13.1	44	46.8	74	63.5
5	15.4	45	47.4	75	64.2
6	17.3	46	47.9	76	64.9
7	18.9	47	48.4	77	65.6
8	20.4	48	48.9	78	66.3
9	21.8	49	49.5	79	67.0
10	23.0	50	50.0	80	67.7
11	24.2	51	50.5	81	68.5
12	25.3	52	51.1	82	69.3
13	26.3	53	51.6	83	70.1
14	27.2	54	52.1	84	70.9
15	28.2	55	52.6	85	71.8
16	29.1	56	53.2	86	72.8
17	29.9	57	53.7	87	73.7
18	30.7	58	54.2	88	74.7
19	31.5	59	54.8	89	75.8
20	32.3	60	55.3	90	77.0
21	33.0	61	55.9	91	78.2
22	33.7	62	56.4	92	79.6
23	34.4	63	57.0	93	81.1
24	35.1	64	57.5	94	82.7
25	35.8	65	58.1	95	84.6
26	36.5	66	58.7	96	86.9
27	37.1	67	59.3	97	89.6
28	37.7	68	59.9	98	93.3
29	38.3	69	60.4	99	99.0
30	39.0	70	61.0		
31	39.6				
32	40.1				
33	40.7				
34	41.3				
35	41.9				
36	42.5				
37	43.0				
38	43.6				
39	44.1				
40	44.7				


14 16
5
18 20
4
17 20
3
10
30
50
70
PERCENTILES

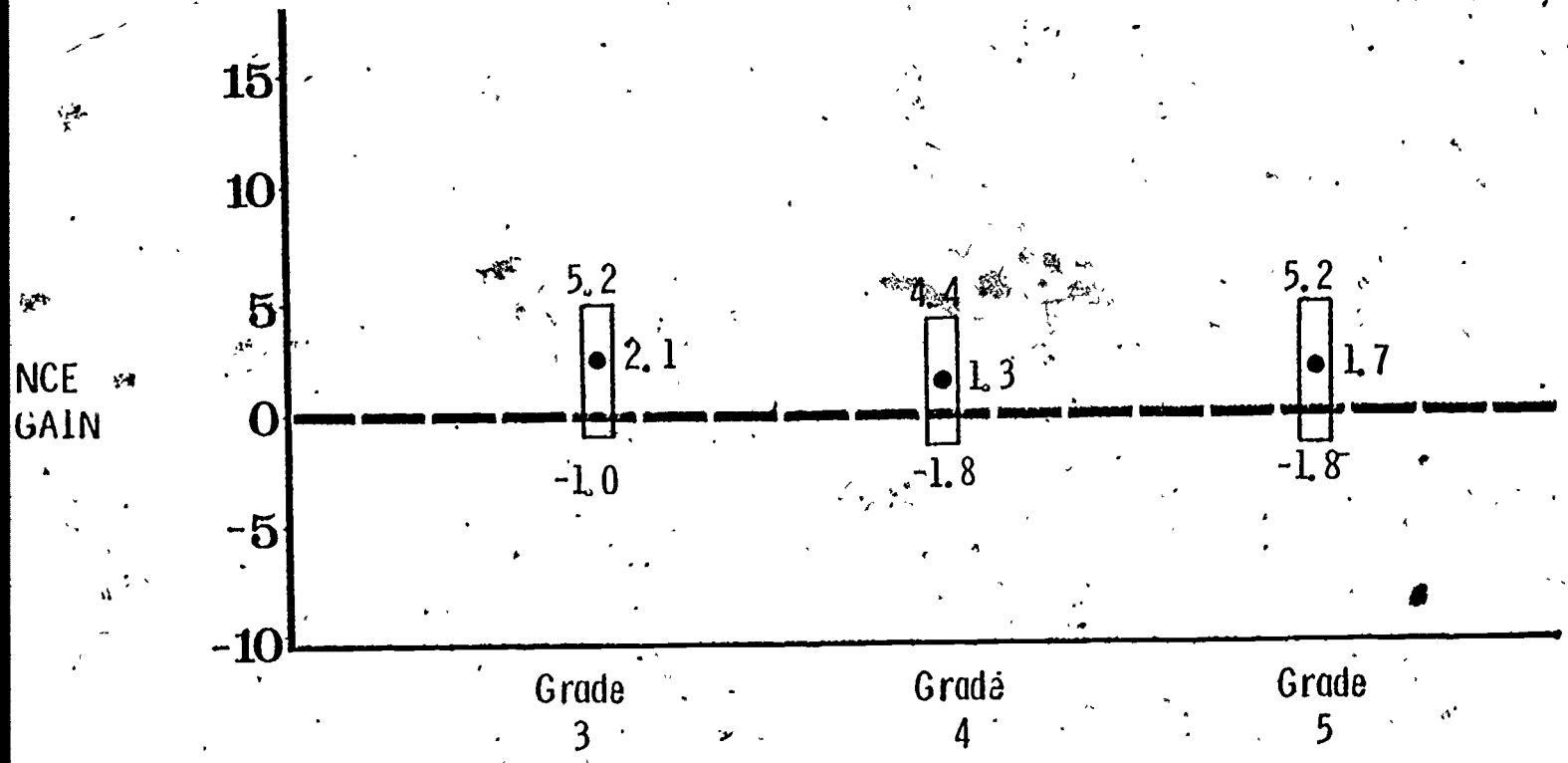

16
14
12
10
8
6
4
2
error in NCEs
5
10
15
20
25
30
45
50
class size

"Give or Take" Table

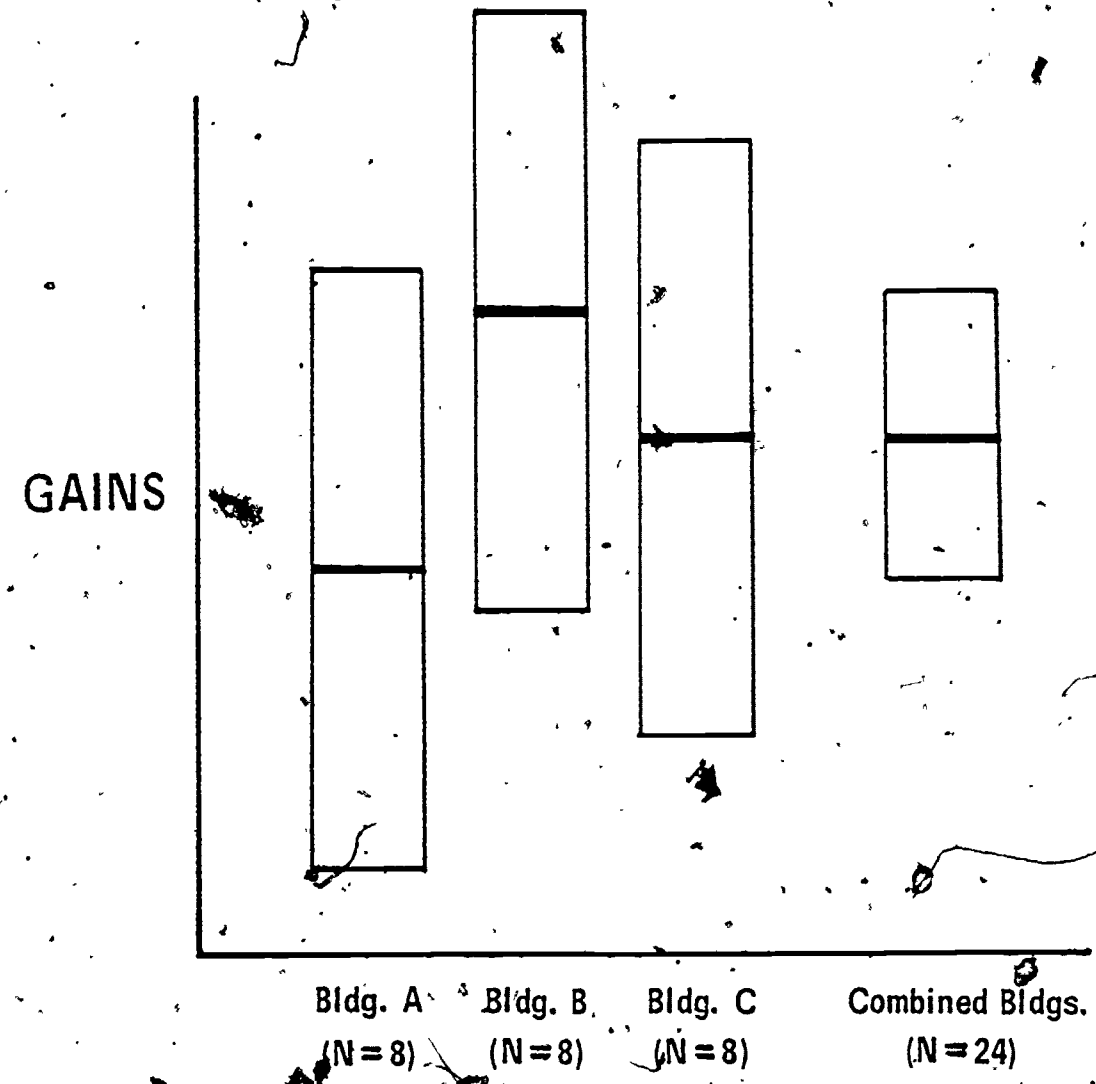
Class Size (N)	Error (NCEs)	Class Size (N)	Error (NCEs)
2	16.0	21	3.6
3	11.3	22	3.5
4	9.2	23	3.4
5	8.0	24-25	3.3
6	7.2	26	3.2
7	6.5	27-28	3.1
8	6.1	29-30	3.0
9	5.7	31-32	2.9
10	5.3	33-34	2.8
11	5.1	35-37	2.7
12	4.8	38-40	2.6
13	4.6	41-43	2.5
14	4.4	44-47	2.4
15	4.3	48-50	2.3
16	4.1		
17	4.0	75	1.9
18	3.9	100	1.6
19	3.8	150	1.6
20	3.7	200	1.1



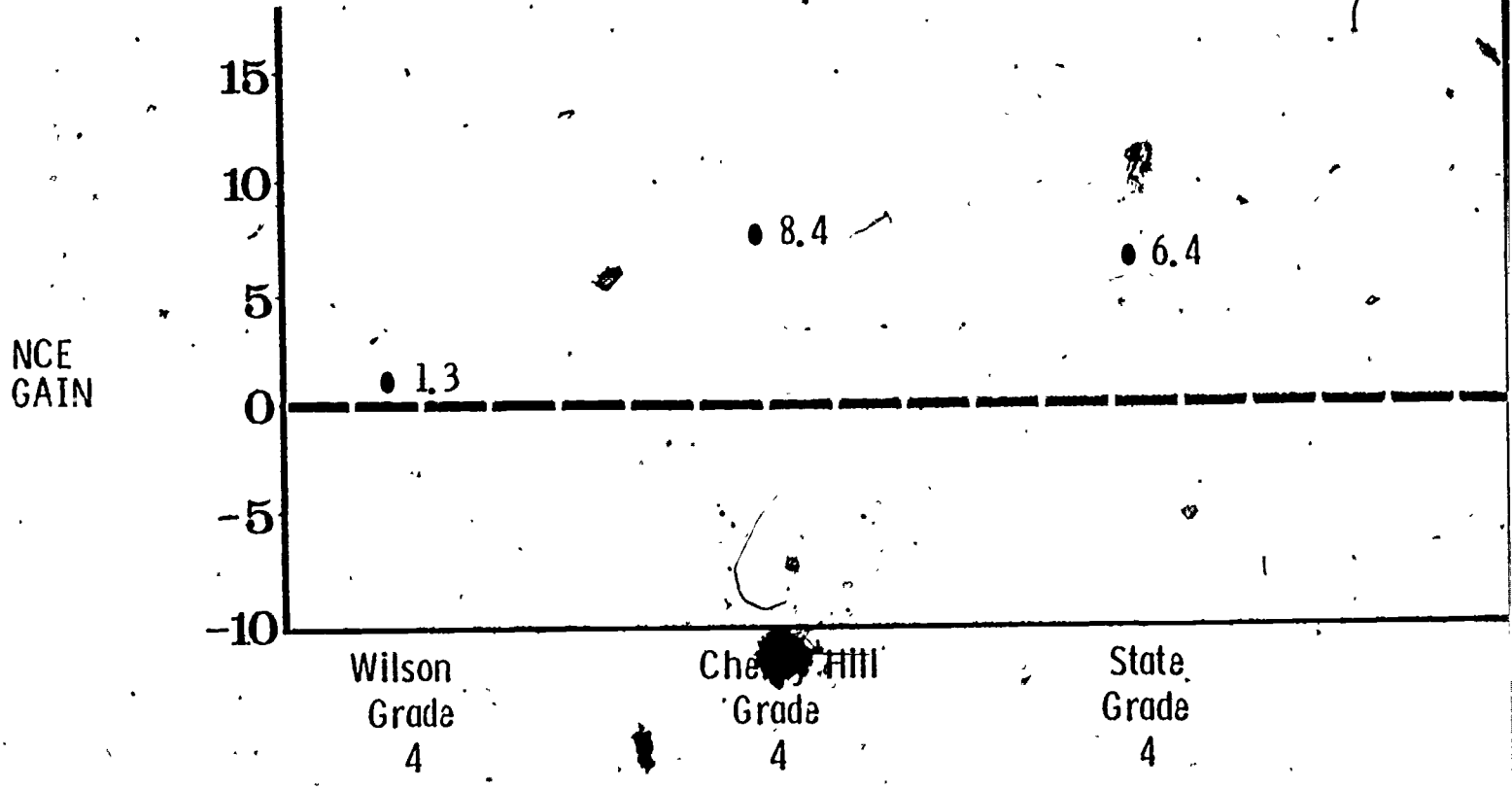
WILSON ELEMENTARY SCHOOL



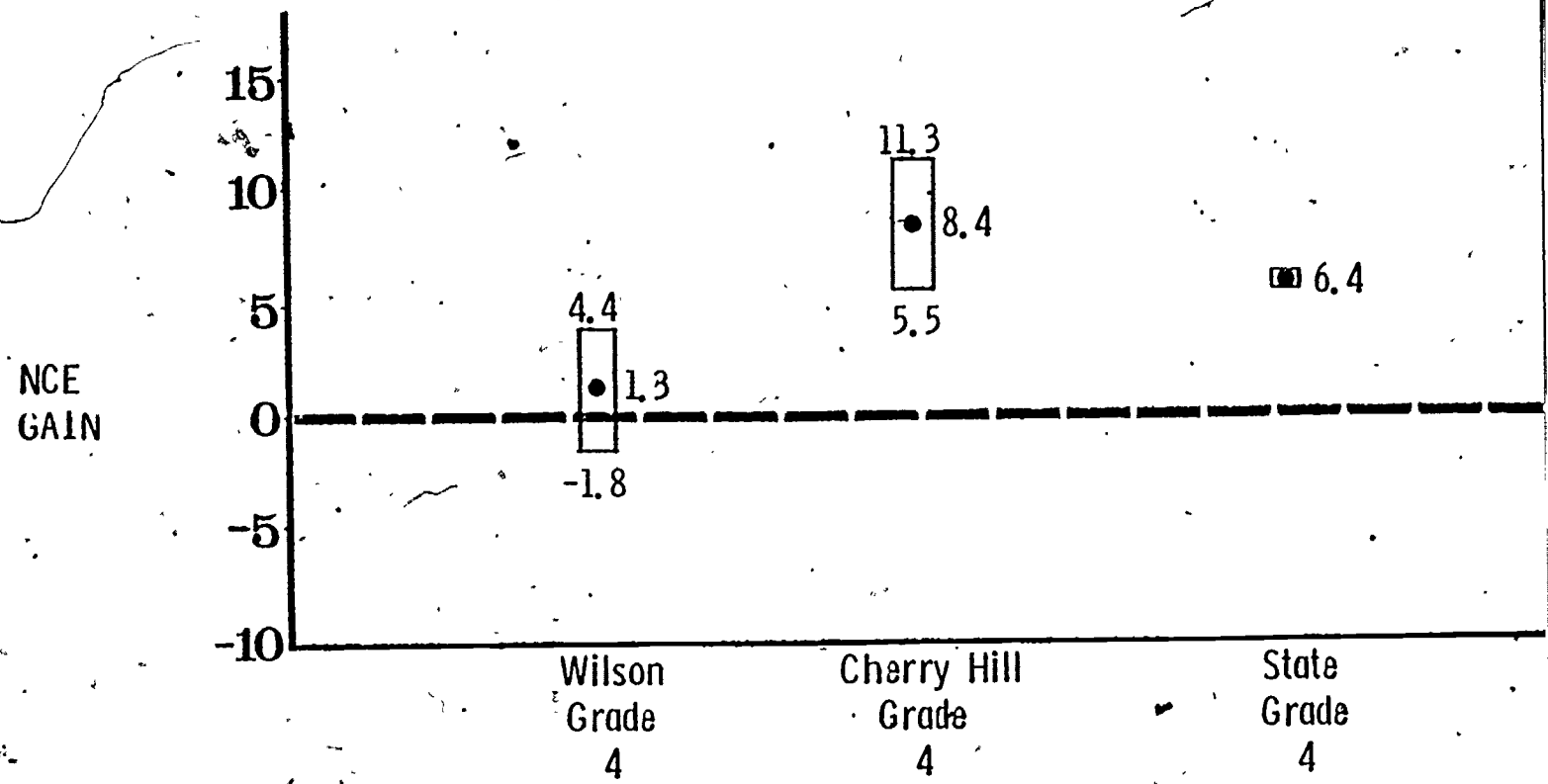
WILSON ELEMENTARY SCHOOL


GAINS
Bldg. A (N=8)
Bldg. B (N=8)
Bldg. C (N=8)
Combined Bldgs. (N=24)

N	NORMAL CURVE EQUIVALENTS (NCEs)		
	Pretest	Posttest	Gain
29	31.6	38.3	6.7
32	34.4	42.8	8.4
30	30.9	35.7	4.8

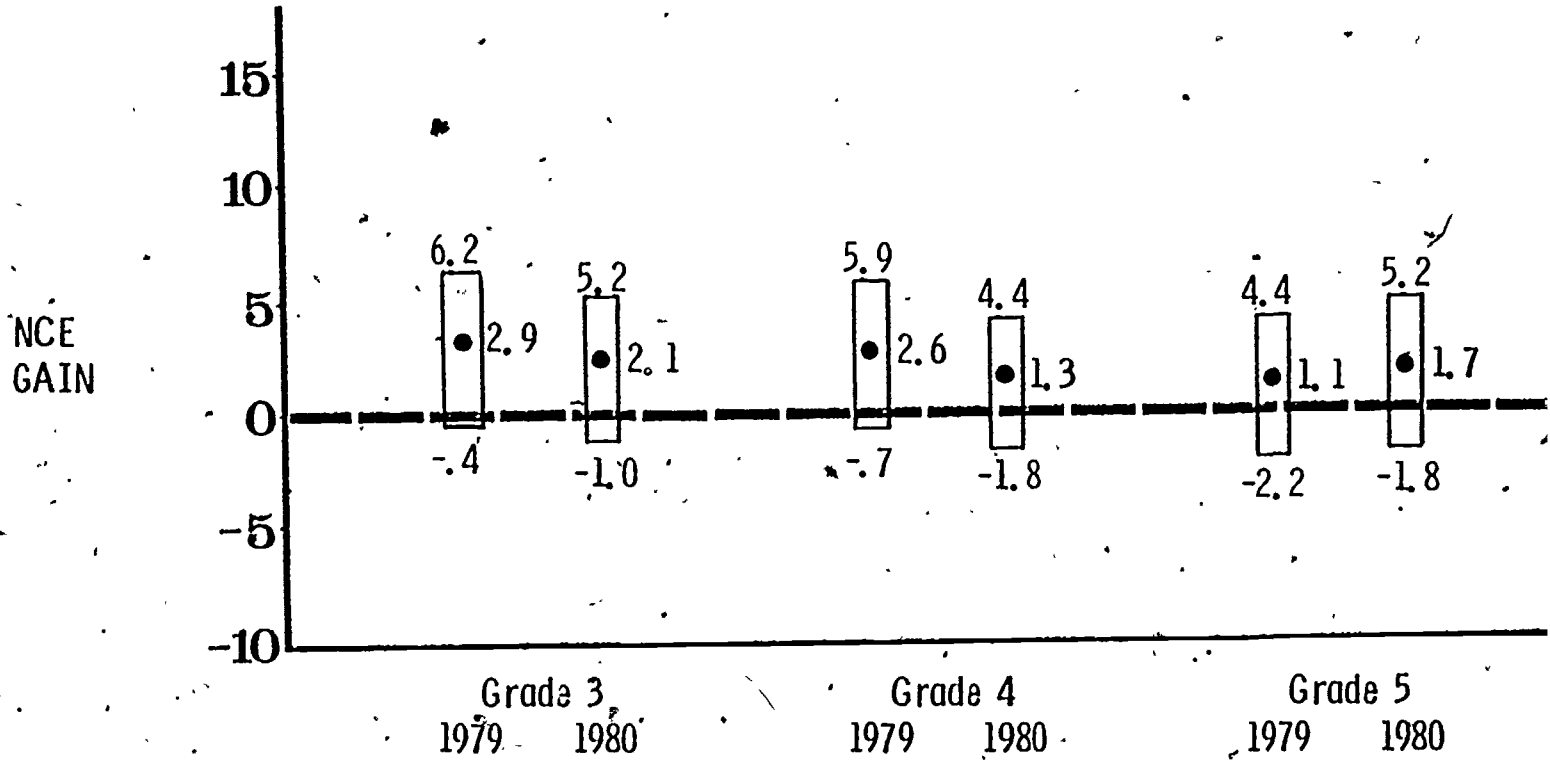


WILSON vs. CHERRY HILL AND STATE



WILSON vs. CHERRY HILL AND STATE ELEMENTARY SCORES

	NORMAL CURVE EQUIVALENTS (NCEs)		
N	Pretest	Posttest	Gain
24	27.8	30.7	2 9
25	29.2	31.8	2.6
24	25.5	26.6	1.1



WILSON 1979 vs. WILSON 1980

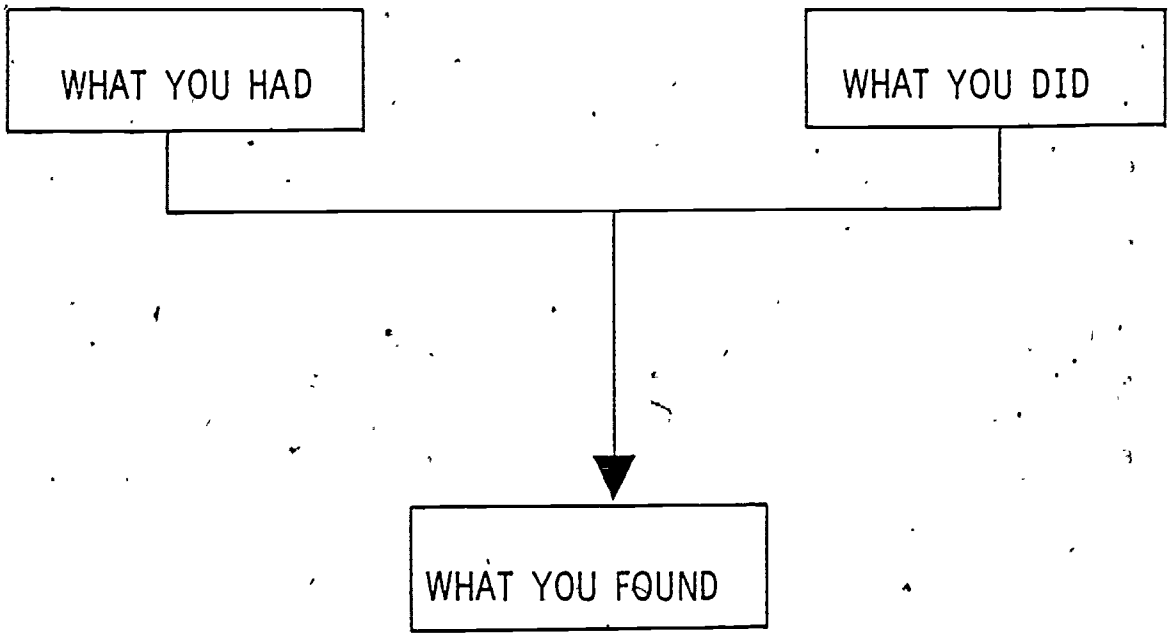
INFORMATION FROM STATE TITLE I REPORT
1980

NCE Gain by Grade, Reading Projects

Grade	NCE Gain
3	7.42
4	6.36
5	5.96

PROGRAM OBJECTIVES
Centennial

- Increase two reading levels
- Improved comprehension (8 NCEs)


WHAT YOU HAD
WHAT YOU DID
WHAT YOU FOUND

INTERPRETING TITLE I EVALUATION RESULTS

Section III: Participant Materials

A. Title I Reading Project Description
Wilson Elementary School

This Year's Achievement Impact Report

B. This Year's Achievement Impact Report
Cherry Hill Elementary School

Last Year's Achievement Impact Report
Wilson Elementary School

This Year's State Title I Achievement Impact Report

Program Objectives for Centennial School District

C. Graphs for Showing Wilson's Results

D. Title I Director's Annual Project Summary

Title I Evaluator's Summary

E. Interpretation Guide for Title I Evaluation Results

A. Title I Reading Project Description
Wilson Elementary School

This Year's Achievement Impact Report

TITLE I READING PROJECT DESCRIPTION
Wilson Elementary School

Background of Project

The Title I reading project in Wilson Elementary School focuses on the development of reading skills in 3rd, 4th and 5th graders. This project has been operating for three years, using the same basic structure.

Instructional Activities and Services

A reading resource lab is staffed by a Title I teacher and one full-time aide, both working .8 FTE. Students are scheduled into the resource room for approximately one-half hour per day, where they work on projects in small groups. Instruction is based on an individual analysis of each student's needs, although the instruction may involve groupings of objectives and students may work on several objectives at the same time. Mastery of specific objectives is determined by a teacher-developed system of measuring performance. In general, ten students work in the room and interact with the teacher on an individual basis.

The Title I teacher coordinates the activities of the resource lab. The teacher's responsibilities include supervision of the aide; consulting with classroom teachers regarding special needs of target students; individual diagnosis; design of instructional strategies to meet needs of target students; direct instruction to target students in the resource lab; and assessment of student progress on objectives.

The aide assists the teacher in any capacity as determined by the teacher.

Evaluation

The Wilson Elementary School based its most recent evaluation on a norm-referenced evaluation model. This evaluation model compares the average score of Title I students to national norms at two points in time. A test administered prior to the start of the program is used to set the expected percentile standing of the Title I students on the posttest. The expected percentile is the average percentile standing of the Title I students at the pretest. If there were no Title I program, the average posttest percentile is expected to equal this value. The difference between the observed posttest standing and the expected posttest standing for the group is the measure of Title I program effect. For the reading program, the reading subtest, Form O, of Level IV (3rd and 4th grade) and Level V (5th grade) of the Comprehensive Achievement Battery was given and the total reading subtest score was used in the evaluation. The test was administered in October and April.

ACHIEVEMENT INFORMATION

A

School Code 7 School Name Wilson Elementary

District Code _____ District Name Centennial

Project Identification Title I Reading Project

Total number of nonpublic school Title I students included in this summary 0

B

SUBJECT MATTER

(X) 1. Reading
( ) 2. Language Arts
( ) 3. Mathematics
( ) 4. Other (explain on reverse side)

EVALUATION MODEL

(X) 1. A1
( ) 2. A2
( ) 3. B1
( ) 4. B2
( ) 5. C1
( ) 6. C2
( ) 7. Alternative

TESTING INTERVAL

( ) 1. Spring to Spring
( ) 2. Fall to Fall
(X) 3. Fall to Spring
( ) 4. Other (explain on reverse side)

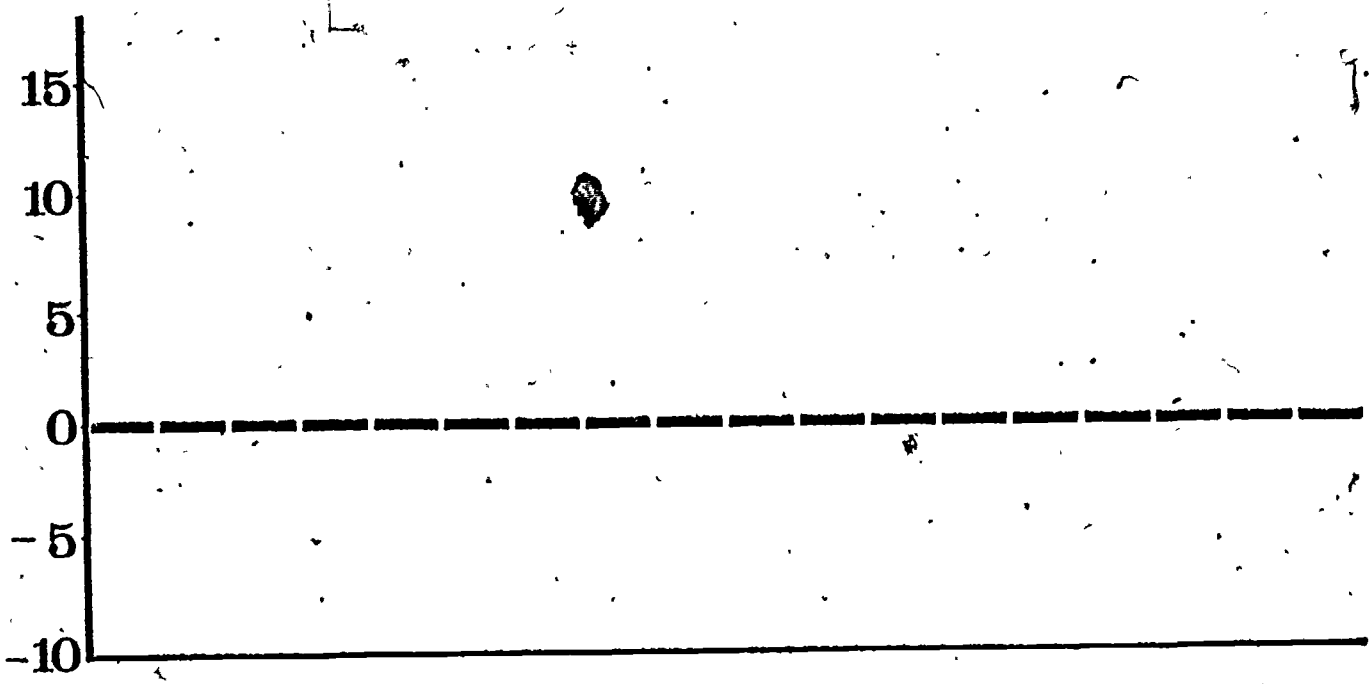
C

	Test Code*	Test Edition *	Date Administered (Month/Day)
PRETEST	1 7	19 7 6	10 / 10
POSTTEST	1 7	19 7 6	4 / 23
NORMED TEST (for A2, B2 & C2 only)		19	/

D

Grade	Project* Setting Code	Hrs/Week Instruction	Project Length in weeks	Student to Instructor Ratio (S:I)	Out-of-Level Testing	Membership	N	NORMAL CURVE EQUIVALENTS (NCEs) Pretest	Posttest	Gain
2				:	Y N					
3	2	2.5	32	10 : 2	Y (N)	30	27	29.9	32.0	2.1
4	2	2.5	32	10 : 2	Y (N)	30	28	30.8	32.1	1.3
5	2	2.5	32	10 : 2	Y (N)	30	22	27.1	28.8	1.7
6				:	Y N					
7				:	Y N					
8				:	Y N					
9				:	Y N					
10				:	Y N					
11				:	Y N					
12				:	Y N					

*SEE BACK OF PAGE FOR CODES


15
10
5
0
−5
−10

B. This Year's Achievement Impact Report
Cherry Hill Elementary School

Last Year's Achievement Impact Report
Wilson Elementary School

This Year's State Title I Achievement Impact Report

Program Objectives for Centennial School District

ACHIEVEMENT INFORMATION

A

School Code _____ School Name Cherry Hill

District Code _____ District Name Centennial

Project Identification Title I Reading Project

Total number of nonpublic school Title I students included in this summary 0

B

SUBJECT MATTER

(X) 1. Reading
( ) 2. Language Arts
( ) 3. Mathematics
( ) 4. Other (explain on reverse side)

EVALUATION MODEL

(X) 1. A1
( ) 2. A2
( ) 3. B1
( ) 4. B2
( ) 5. C1
( ) 6. C2
( ) 7. Alternative

TESTING INTERVAL

( ) 1. Spring to Spring
( ) 2. Fall to Fall
(X) 3. Fall to Spring
( ) 4. Other (explain on reverse side)

C

	Test Code*	Test Edition *	Date Administered (Month/Day)
PRETEST	1 7	19 7 6	10 / 10
POSTTEST	1 7	19 7 6	4 / 23
NORMED TEST (for A2, B2 & C2 only)		19	/

D

Grade	Project* Setting Code	Hrs/Week Instruction	Project Length in weeks	Student to Instructor Ratio (S:I)	Out-of-Level Testing	Membership	N	NORMAL CURVE EQUIVALENTS (NCEs) Pretest	Posttest	Gain
2				:	Y N					
3	2	2.5	32	10 : 2	Y (N)	35	29	31.6	38.3	6.7
4	2	2.5	32	10 : 2	Y (N)	35	32	34.4	42.8	8.4
5	2	2.5	32	10 : 2	Y (N)	35	30	30.9	35.7	4.8
6				:	Y N					
7				:	Y N					
8				:	Y N					
9				:	Y N					
10				:	Y N					
11				:	Y N					
12				:	Y N					

*SEE BACK OF PAGE FOR CODES

ACHIEVEMENT INFORMATION

A

School Code _____ School Name Wilson Elementary

District Code _____ District Name Centennial

Project Identification Title I Reading

Total number of nonpublic school Title I students included in this summary 0

B

SUBJECT MATTER

(x) 1. Reading
( ) 2. Language Arts
( ) 3. Mathematics
( ) 4. Other (explain on reverse side)

EVALUATION MODEL

(x) 1. A1
( ) 2. A2
( ) 3. B1
( ) 4. B2
( ) 5. C1
( ) 6. C2
( ) 7. Alternative

TESTING INTERVAL

( ) 1. Spring to Spring
( ) 2. Fall to Fall
(x) 3. Fall to Spring
( ) 4. Other (explain on reverse side)

C

	Test Code*	Test Edition *	Date Administered (Month/Day)
PRETEST	1 7	19 7 6	10 / 8
POSTTEST	1 7	19 7 6	4 / 20
NORMED TEST (for A2, B2 & C2 only)		19	/

D

Grade	Project* Setting Code	Hrs/Week Instruction	Project Length in weeks	Student to Instructor Ratio (S:I)	Out-of-Level Testing	Membership	N	NCE Pretest	NCE Posttest	NCE Gain
2				:	Y N					
3	2	2.5	32	10 : 2	Y (N)	30	24	27.8	30.7	2.9
4	2	2.5	32	10 : 2	Y (N)	30	25	29.2	31.8	2.6
5	2	2.5	32	10 : 2	Y (N)	30	24	25.5	26.6	1.1
6				:	Y N					
7				:	Y N					
8				:	Y N					
9				:	Y N					
10				:	Y N					
11				:	Y N					
12				:	Y N					

*SEE BACK OF PAGE FOR CODES

THIS YEAR'S REPORT FROM THE STATE

NCE Gain by Grade, Reading Projects

Grade	NCE Gain
3	7.42
4	6.36
5	5.96

Program Objectives for Centennial Title I Reading Project

As a result of program participation, target students will demonstrate improved reading and language skills:

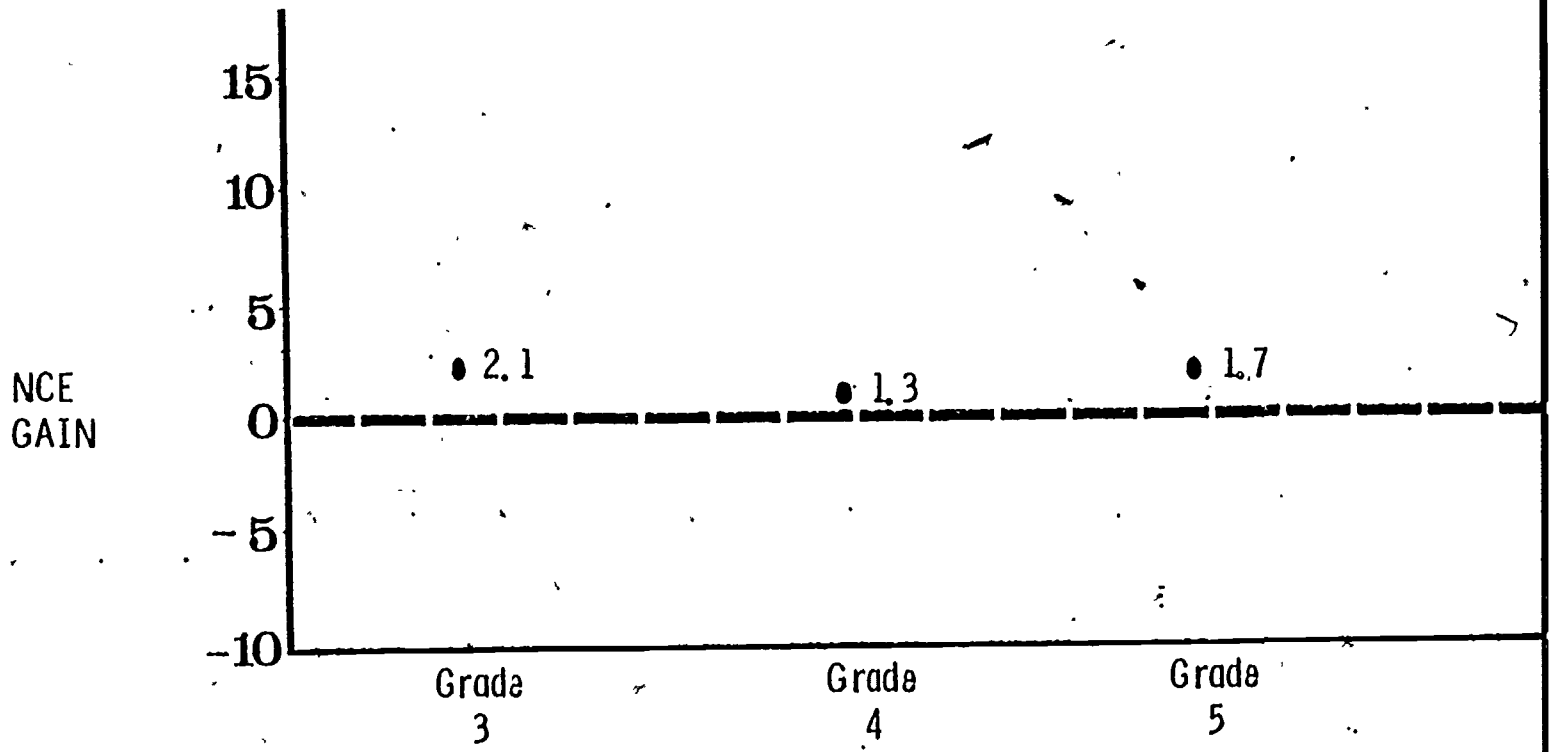
1. Target students will show an increase in overall reading skill as demonstrated by an increase of two reading book levels during the project year.

2. Target students will show improved reading comprehension skills as demonstrated by an average gain of eight standard scores (NCEs) above their expected score without Title I. The reading comprehension subtest of the Comprehensive Achievement Battery will be used to measure reading skills.

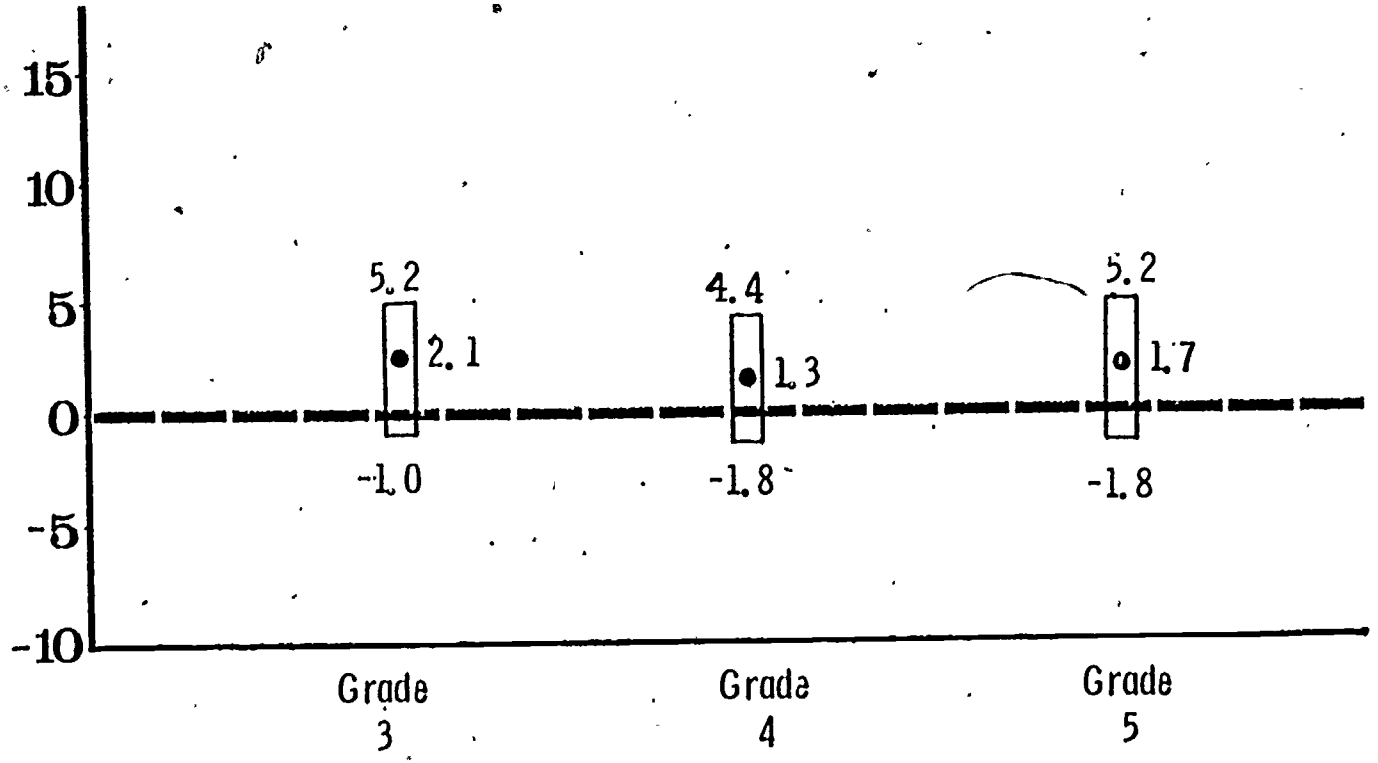
Individual students in the target population will receive diagnosis and prescriptive instruction focused on each child's needs.

3091D
1/81

C. Graphs for Showing Wilson's Results

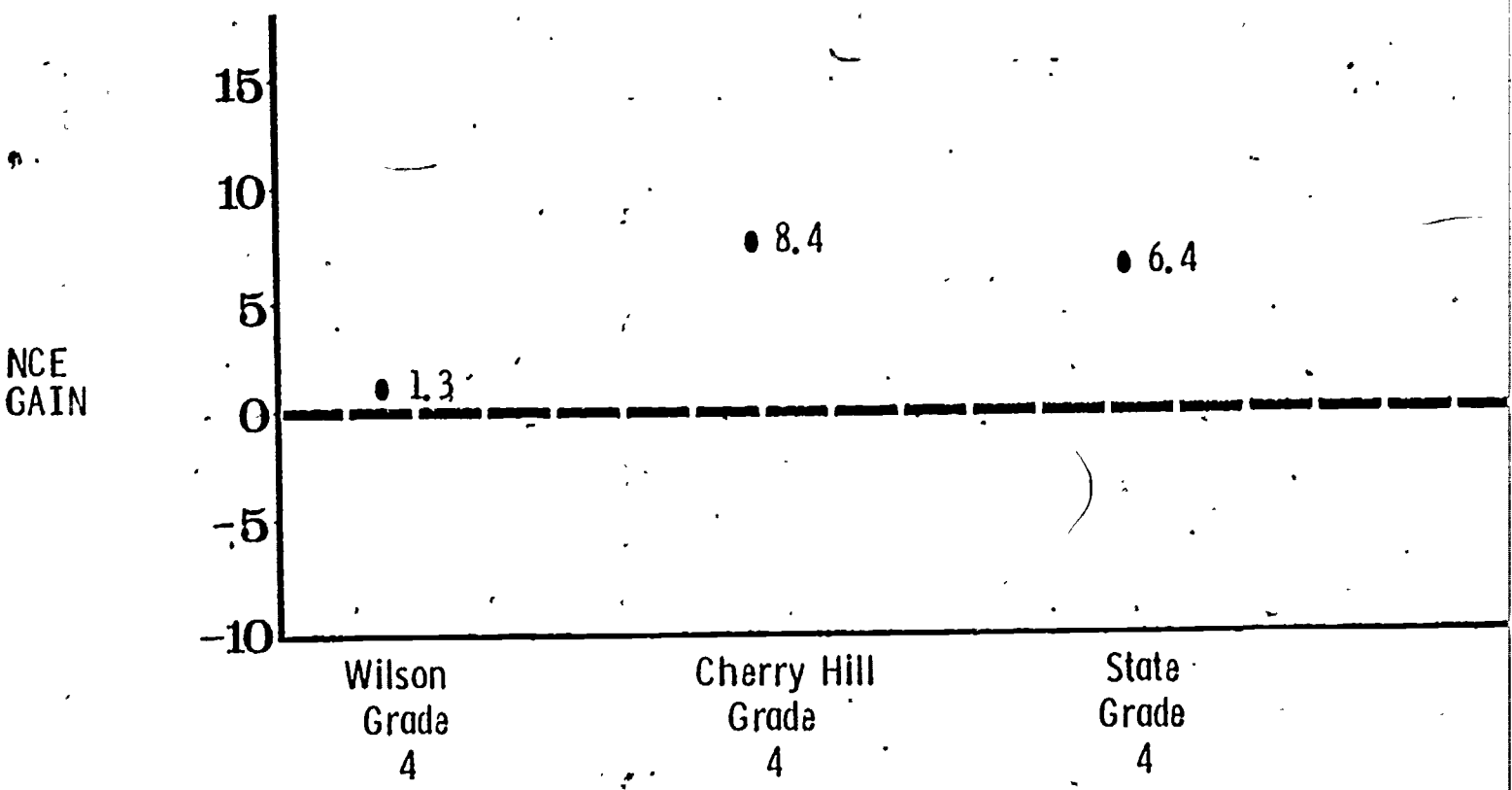


WILSON ELEMENTARY SCHOOL

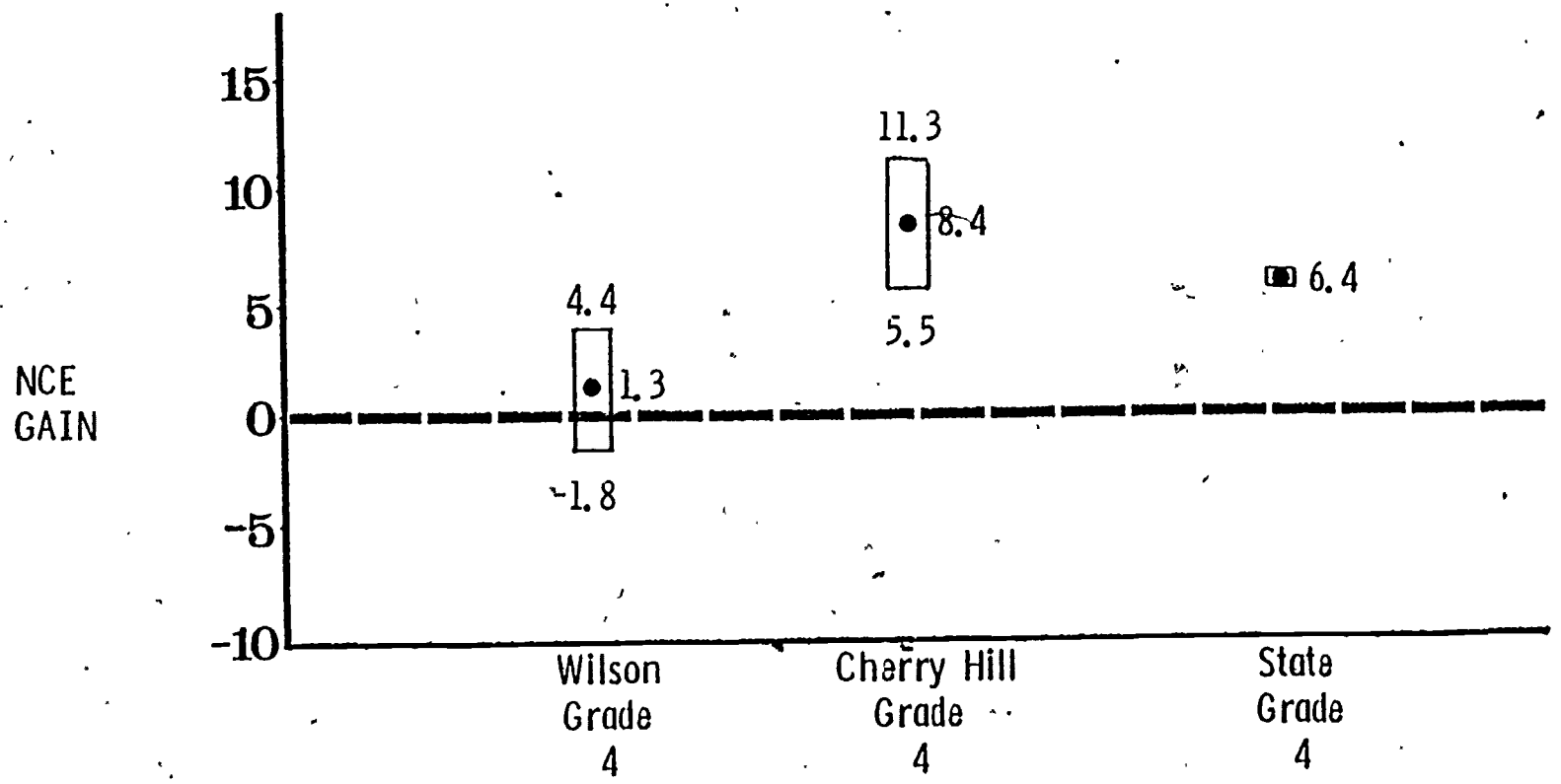


WILSON ELEMENTARY SCHOOL

ANSWERS

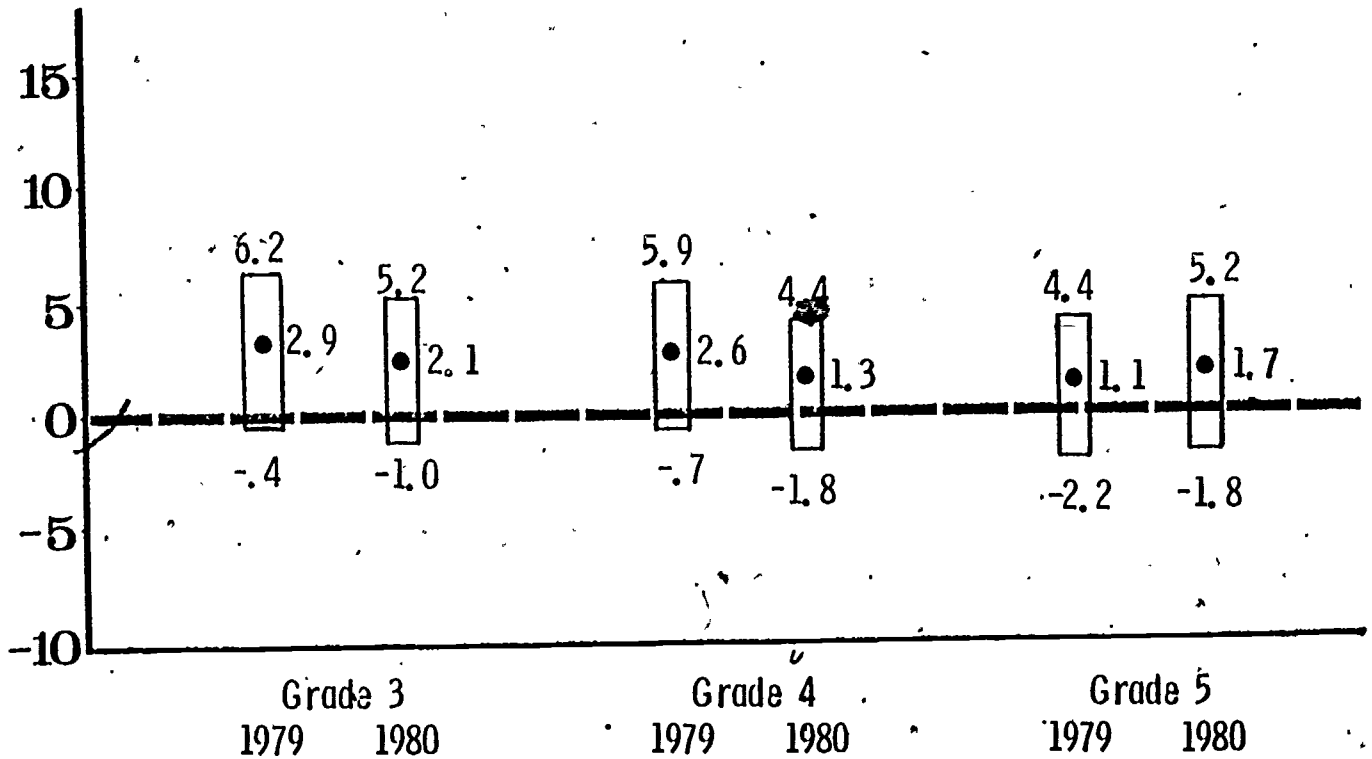


WILSON vs. CHERRY HILL AND STATE



WILSON vs. CHERRY HILL AND STATE ELEMENTARY SCORES

ANSWERS



WILSON 1979 vs. WILSON 1980

ANSWERS

D. Title I Director's Annual Project Summary

Title I Evaluator's Summary

TITLE I DIRECTOR'S ANNUAL PROJECT SUMMARY
Wilson Elementary School Title I Reading Project

The Title I reading project this year has been moving fairly well. I feel that students made a great deal of progress, although the program evaluation indicated that we did not meet all of our goals.

Overall Program

Scheduling students into the resource room for half an hour a day seems to work well. As always, we noted a good deal of absenteeism on Mondays, but attendance was high the rest of the week.

Several of the new classroom teachers made little use of the resource room and our special program. Several of them seemed to prefer dealing with Title I students in their own classrooms. But we are working on better procedures for identifying all Title I-eligible students next year.

In my dual role as project director and teacher, I have been able to observe the program both from the instructional and from the administrative point of view, and I feel very good about our progress. The aide and I worked well together and, as a team, were able to keep students interested and involved in the program. As part of my record-keeping, I kept track of how we both spent our time over the past year. A table summarizing our activities is included. I felt that we spent somewhat longer than necessary in recording student achievement and on providing direction (rather than instructional help) to students. Several recommendations are included to avoid these problems in the future.

Tasks	Percentage of Time on Tasks	
	Teacher	Aide
Administration	11	2
Testing students—formal tests	4	8
Checking mastery with informal tests	22	26
Developing new projects or materials	13	4
Correcting assignments	12	19
Giving assignments	13	5
Monitoring student work on projects	8	21
Talking with students	7	7
Talking with each other	8	8
Talking with parents	2	0

Student Mastery of Objectives

While our project's performance in terms of NCE gains is somewhat disappointing, we have made observable progress with the students. During the past year, as students mastered specific objectives, we marked the objectives on a student achievement record. At the end of each month, my aide records the number of Title I students mastering each objective. These monthly charts were used to examine student progress on objectives.

The program objectives cover four different skill areas: phonetic analysis, structural analysis, vocabulary and reading comprehension. About half of the students entering the program had already mastered at least 73 percent of the phonetic analysis objectives, 58 percent of the structural analysis objectives, 36 percent of the vocabulary objectives and 17 percent of the reading comprehension objectives. By the end of the year, 80 percent of the students had mastered all of the phonetics and vocabulary objectives, 84 percent the structural analysis ones and 44 percent the comprehension ones. These figures seem to indicate that progress is made on all of the objectives of the program.

Recommendations

Before school begins in the fall I would like to be able to spend some time preparing individualized packets for each student who will be in the resource room. If each student received a packet detailing instructions for several alternative activities for each objective, explaining and encouraging the use of different learning options would be greatly facilitated.

We also need to spend some time with our recordkeeping systems, so that when student-teacher and parent-teacher conferences are held it will be easier to access the relevant information. I recommend sending my aide and myself to the recordkeeping clinic sponsored by the publishers of the Special Readers Book Series we are using.

Title I Evaluator's Summary

This report summarizes an evaluation of the Wilson Elementary Reading Program over the past year.

Impact Evaluation. Given the size of Centennial School District and the lack of an adequate comparison group of students at Wilson, Model A was used to evaluate the Title I program. In conducting the study, Model A guidelines were followed. Students were tested at the fall and spring empirical norm dates of the Comprehensive Achievement Battery, student selection was independent of pretest scores, all students were tested within a short timespan, and the test administrators followed the publisher's directions.

The impact data showed October to April gains of 2.1 NCEs for the third graders, 1.3 NCEs for fourth graders and 1.7 NCEs for fifth graders. These gains mean that as a group, third graders moved from the 17.0 percentile to the 21.0 percentile, fourth graders from the 18.1 to the 19.8 percentile, and fifth graders from the 13.8 to the 15.7 percentile. The impact data are based upon the total reading subtest of the Comprehensive Achievement Battery (CAB).

It should be stated that any NCE gains greater than 0 indicate some program effect. A comparison of Wilson gains with the stated program objectives shows a much smaller gain than expected, however.

Even a comparison to the gains achieved by students at Cherry Hill Elementary School, a similar school with a similar program, shows that the Wilson gains are probably not as large as desired. A comparison of the reading programs at Cherry Hill and Wilson may indicate some possible explanations for the differences in achievement. For example, the proportion of objectives in each skill area differs dramatically for the two schools. The areas of objectives and number of objectives in each area are listed below.

Table 1

Area	No. of Objectives	
	Wilson	Cherry Hill
Phonetic analysis	22	22
Structural analysis	12	12
Vocabulary	11	11
Reading comprehension	6	21

Given the differences in the level of objective for the two programs, April test scores for Title I students in both schools were computed for all reading subtests of the CAB. A description of the test content is attached.

A further analysis of the Wilson program was based upon student mastery of objectives. Overall student mastery is summarized in the Title I director's report. Further study of student progress rates also showed that most student mastery of phonic analysis and structural analysis objectives occurred in the first month of the program. Little progress was seen during the last half of the project. This would seem to indicate some problem in the program structure.

Table 2
Percentile Ranks of Title I Students on the
Comprehensive Achievement Battery
(April Testing)

Elementary		Subtests				Total	
Grade	School	Phonetics	Word Analysis	Vocabulary	Comprehension	%ile	NCE
3rd	Cherry Hill	38	35	25	26	29	38.3
	Wilson	37	36	22	14	20	32.0
4th	Cherry Hill	39	42	31	35	37	42.8
	Wilson	45	38	25	15	20	32.1
5th	Cherry Hill	37	30	28	22	25	35.7
	Wilson	42	30	22	10	16	28.8

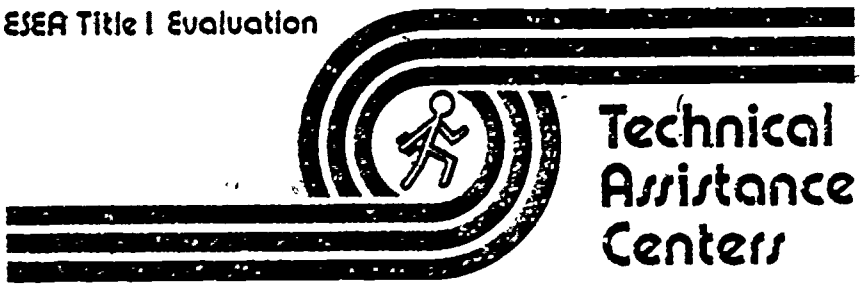
9/80
1148D

The Comprehensive Achievement Battery provided a content description of each subtest. You copied the description for the four reading subtests which made up the Total Reading score used to evaluate the program. For each general content area, the number of test items measuring the skill was listed.

Skill Area	Number of Items Level IV (3rd&4th grade)	Level V (5th grade
Phonetics Skills Subtest		
1. Matching Oral Sounds: Find words containing the same sound in words read aloud by the teacher.	10	--
2. Matching sounds in printed words: Find words containing the same sounds as words printed in the test manual.	10	10
Word Analysis Skills Subtest		
1. Verbs: Select the appropriate verb endings.	7	7
2. Adjectives: Select the appropriate adjective endings.	7	7
3. Combinations: Make contractions and compound words.	6	6
Vocabulary Skills Subtest		
1. Single Meaning: Select the correct definition of a word that has only one meaning.	15	10
2. Multiple Meaning: Select the correct definition of a word that has more than one meaning.	15	10
Reading Comprehension Skills Subtest		
1. Restate Material: Recognize a restatement of material that is explicity stated in the passage; select a specific detail that is stated in the passage.	10	10
2. Sequence and Summarize: Select the proper sequence of ideas or action in the passage; select main ideas or choose an appropriate title for the passage.	10	10
3. Draw Inferences: Recognize material that is implied but not specifically stated in the passage; recognize a character's motivations and emotions; select probable reasons for actions.	10	10
4. Apply to New Situations: Recognize a valid example of something stated in the passage; choose a likely outcome with one variable changed; select correct applications of the information in the passage to a new situation.	--	10
5. Logical Relationships: Recognize the relation between premises or sections of a passage; recognize validity of procedures and variables in science and social studies passages; distinguish between fact and opinion; choose statements or examples of technique and point of view.	--	10

E. Interpretation Guide for Title I Evaluation Results

ESEA Title I Evaluation



INTERPRETATION GUIDE FOR TITLE I EVALUATION RESULTS

TABLE OF CONTENTS

INTRODUCTION

So you've completed the end-of-the-year report for your Title I program. You've described your program, averaged pretest and posttest scores, and calculated NCE gains. There's a feeling of relief yet also one of curiosity. Surely there must be more to evaluation than completing forms. What do these results tell you? Did your program work? Are these numbers in front of you acceptable? Are they low? High? Can you use them to tell you anything about your program? Are there parts of the program that should be revised?

What this guide will do

This guide is designed to take you from simple explanations of what an NCE gain means to more complicated concerns such as where to look in the program for the strengths and weaknesses. To do this, the guide is divided into three major sections.

Section One deals with interpretations of NCEs and NCE gains. You will find explanations as well as alternative ways of describing gains. Also, problems of interpreting gains for small class sizes will be discussed.

Section Two builds on these initial interpretations by comparing these gains to different standards. These standards include last year's results, other projects' results, state averages, or objectives set by your program.

Section Three discusses aspects of your program that could be causing the project outcomes you have observed. You can use this information to diagnose program strengths and weaknesses.

What this guide won't do

Each section of this guide progressively adds to the interpretation of your results. Since Title I programs differ widely in their characteristics, providing you with absolute results would be impossible. The final interpretation can only be made by you. This guide will not, therefore, provide you with the answers to all your specific project questions. Hopefully, it will provide you with the necessary background information to discover those answers on your own.

What you need before using this guide

To use this guide, you must be able to assume that your data are a valid reflection of your program's impact. This means that you have:

- Used a test which is sensitive enough to detect the effects of the project
- Administered the test according to publisher's instructions

- Controlled for any conditions that could have invalidated the test results (such as room conditions, anxiety, extraneous noise, etc.)

- Followed guidelines for particular evaluation models (such as testing near norm dates, functional-level testing, etc.)

- Checked the data according to the Title I Evaluation Technical Standards--representativeness, reliable and valid instruments and procedures, minimization of error, and valid assessment.

It is next to impossible to have perfectly valid data to interpret. Errors will creep in somewhere. However, to the extent that you have followed these guidelines, you can be more confident in using these data for decisions about your program.

Where to begin in the guide

As mentioned earlier, the guide is divided into three sections, each more advanced than the one before it. You may not have to start at the beginning and may want to begin work somewhere later in the guide. You will probably find it worthwhile to at least leaf through some earlier sections to assure yourself of that familiarity.

A running example

To illustrate the points made throughout the manual, a running example is provided based on the experiences of Centennial School District in interpreting their Title I results. Each part of the example is highlighted with a box.

Example--Project Description

During the 1980-81 school year, Centennial School District implemented Title I programs in elementary reading and middle school language arts. A Title I coordinator was assigned to monitor the instruction and to conduct the evaluation for these programs.

The reading program was implemented in third, fourth, and fifth grades in both Wilson Elementary and Cherry Hill Elementary. Students worked in a resource room staffed by a teacher and an aide for about 30 minutes each day.

The language arts program at Centennial Junior High included seventh and eighth grade students. Remedial instruction was provided by the classroom teacher while the rest of the class worked on other assignments.

SECTION ONE

WHAT ARE NCE GAINS?

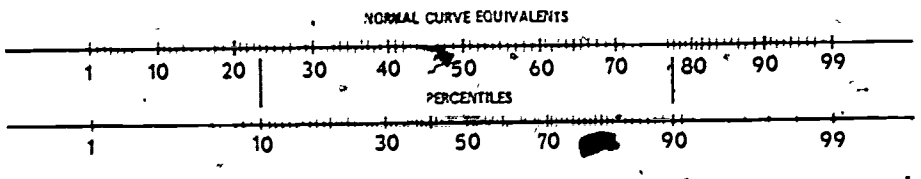
What are NCEs?

NCEs (Normal Curve Equivalents) are the scores used to report Title I program evaluation. While there are already many types of test scores available, NCEs were developed to avoid problems which occur when using other scores for program evaluation. Let's compare NCE scores to percentile ranks with which you are familiar.

A percentile rank compares a student's performance on a test with the performance of students in the norming sample for that test, yielding the percentage of norm group students who fell below a particular score. An NCE score also represents a student's level of achievement compared with students in the norm group, but using different units.

To evaluate the effectiveness of your program you need to average the scores of your Title I students to determine the performance of the whole group. To do this, you must use a score that has equal intervals. You should average NCEs rather than percentiles.

The NCE scale and the percentile scale are matched at the 1st, 50th, and 99th points. This can be seen in the figure below which presents a comparison of the two scales.



Notice that the distances between points on the percentile scale are not equal. At the upper and lower ends of the scale, the points are farther apart than at the middle of the scale. On the NCE scale, the distance between points is the same. For this reason the NCE scale is referred to as an equal interval scale.

You can also compute the average using expanded standard scores since these scores have equal intervals, too. (Note that the publisher of your test may use a different name for these scores--standard scores, expanded scale score, growth scale value.) These scores must generally be used if you have tested your students at a level below that recommended by the test publisher.

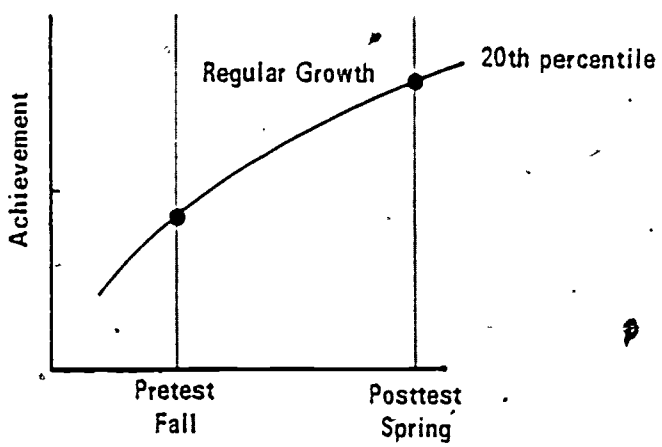
The advantage of NCEs over expanded scores is that NCEs mean the same thing regardless of the test used or the grade level. Expanded scores on the other hand, are not comparable from one test to another or from one grade to another.

In summary, NCEs have two primary advantages for program evaluation over other types of scores--they can be averaged and they are comparable across grade levels and across tests.

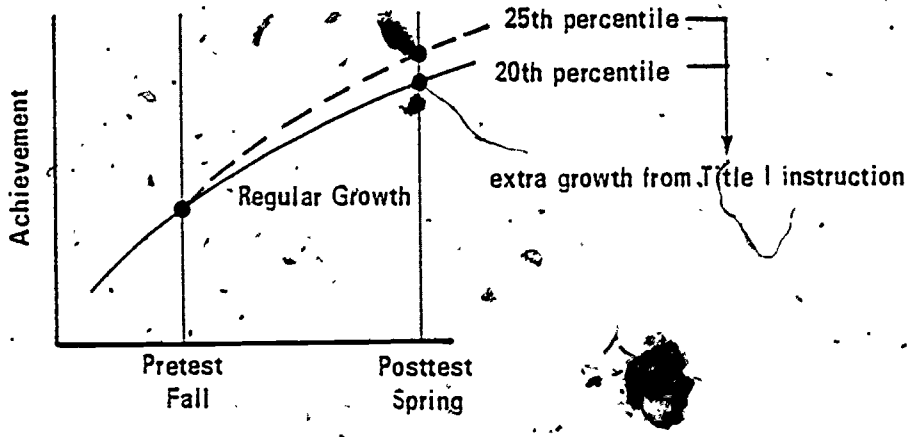
What is the meaning of NCE gains?

Title I students are pretested early in the program and posttested late in the program to determine how much they have grown in achievement. This growth is due to regular classroom instruction, maturation, and home influences as well as the supplementary Title I instruction. To isolate how much of this growth is due just to your program, the Title I evaluation compares the performance of Title I students with similar students who did not receive Title I instruction. Consider how this is done using Model A, the norm-referenced evaluation model.

Model A assumes that the norms for a test reflect the growth expected with regular instruction. Students grow in achievement at different rates, the better students learning faster than the more disadvantaged students. On the average though, a student tends to remain at the same rank relative to his peers. Thus, a group of students whose pretest was at the 20th percentile would be expected to have a posttest average at the 20th percentile.



Title I students, however, have hopefully learned more than is expected for similar students in the regular classroom. Any positive difference found between the average posttest and the average pretest for the Title I group is the gain that can be attributed to Title I instruction. This gain is an indication that extra growth has occurred over and above the regular growth which would have been expected with only regular classroom instruction.



Since NCEs are used instead of percentiles to determine the impact of Title I instruction, the NCE gain for a group of Title I students is the difference between the average NCE on the posttest and the average NCE on the pretest.

NCE GAIN = NCE POSTTEST − NCE PRETEST

The size of NCE gain you should expect may vary depending on the grade level of your students, the subject matter area served, and the nature of your program. It may be said, however, that moderate gains are 5-10 NCEs. Gains of more than 20 NCEs are unusually high and may reflect errors in the way the evaluation was implemented.

What about a gain of zero NCEs? A common misconception concerning a zero gain is that it indicates "no growth at all" has occurred for the group of Title I students. This is incorrect. With a zero NCE gain, the achievement level of your Title I group has increased from pretest to posttest, but this increase in achievement level would have been expected with just regular classroom instruction.

A negative gain indicates that your students did not grow as much as low achieving students in other schools. Alternatively it might suggest that the testing or evaluation were not done properly.

Example--NCE Gains

To evaluate Centennial's Elementary Title I program, a reading test that seemed to match the curriculum was administered fall and spring. The following results were obtained for Wilson Elementary:

Grade	N	Pretest	Posttest	NCE Gain
3	27	29.9	32.0	2.1
4	28	30.8	32.1	1.3
5	22	27.1	28.8	1.7

While there were positive gains at all grades, the gains were small.

How can project gains be more easily explained to others?

When reporting to local audiences such as parents or administrators, you can report the results of your evaluation in ways which are more familiar to these groups. NCEs are useful for computing the group average and for aggregating across buildings or grades but are probably unfamiliar to these audiences. Some alternative methods include: converting NCE averages to more familiar types of scores, graphing NCE gains to emphasize trends, and describing the objectives achieved by the students in the program.

Converting NCE averages. Once you have computed the group's pretest and posttest average NCEs, the average can be converted to other scores. For example, you can express the extra growth of your students due to Title I as an improvement in their percentile rank. While percentiles should not be averaged, they can be used for local reporting.

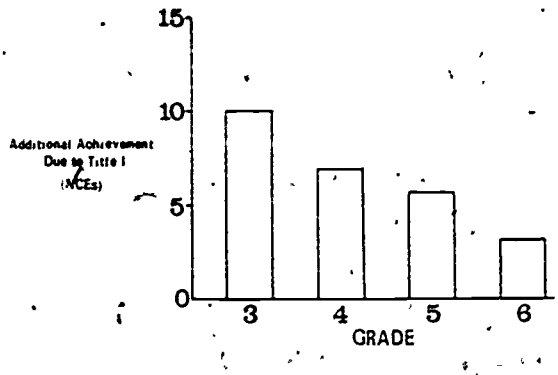
To express your gain as a change in percentile rank, simply convert the pretest and posttest average NCEs to percentiles working backwards in the percentile to the NCE conversion table provided below.

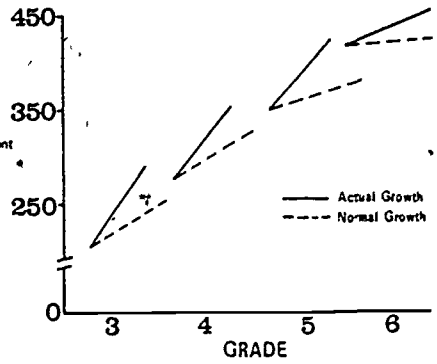
Percentile to NCE Conversion Table

%	NCE	%	NCE	%	NCE	%	NCE	%	NCE
1	1.0	11	24.2	21	33.0	31	39.6	41	45.2
2	6.7	12	25.3	22	33.7	32	40.1	42	45.8
3	10.4	13	26.3	23	34.4	33	40.7	43	46.3
4	13.1	14	27.2	24	35.1	34	41.3	44	46.8
5	15.4	15	28.2	25	35.8	35	41.9	45	47.4
6	17.3	16	29.1	26	36.5	36	42.5	46	47.9
7	18.9	17	29.9	27	37.1	37	43.0	47	48.4
8	20.4	18	30.7	28	37.7	38	43.6	48	48.9
9	21.8	19	31.5	29	38.3	39	44.1	49	49.5
10	23.0	20	32.3	30	39.0	40	44.7	50	50.0

If the pretest and posttest NCEs were 33 and 39, the corresponding percentiles would be 21 and 30. Thus at posttest time, the average Title I student in your program did better than 30 percent of the students in the national norm group, and 9 percentile points better than at the pretest.

Graphing gains. A good graph can draw the readers' attention to trends and basic concepts even if they do not fully understand the underlying scale. A graph of NCE gains might focus the readers' attention on the additional growth due to Title I. A graph of the average pretest and posttest expanded standard scores could be used to compare expected growth (group maintains same percentile) with actual growth. The following graphs show gains graphed using NCEs and expanded standard scores.





Showing content mastery. Rather than talking in generalities about achievement or growth, it may be helpful to focus on the specific skills or objectives learned. If you have good records on the objectives mastered by students, you can talk about the percent of students mastering a set of objectives or the percent of objectives mastered. If you have used the test publisher's scoring services, you probably have an item analysis keyed to objectives that could be used. When hand scoring, you could create your own item analysis by determining the number of items each student got right for clusters of items measuring the same objective.

Are project gains accurate?

While test scores are generally a good indicator of a student's achievement, it is not possible to obtain a perfect measure of achievement. As you know, there are many sources of measurement error which can affect the test score a student obtains. These include fluctuations in the student's mood or motivation, momentary distractions during the test, guessing, and familiarity with the item format. Because of these factors, the student's obtained score at one time will differ--sometimes higher, sometimes lower--from the test score obtained at another time. That is, the score tells you roughly what the student's true achievement level is, but not exactly.

When project gains are computed, the observed gain is affected by errors from both the pretest and posttest. By averaging across all the students in the Title I class, however, this error is reduced since the various sources of error tend to cancel each other. The gains based on only a few students can vary greatly depending on the particular sample of students included. Gains based on many students will be much more accurate.

Since you generally want to know how effective your program is with Title I students in general--not just with this year's group--there is another type of error that affects your results. By the luck of the draw, one year you might have a good group that would make your program look better than it should or a bad group that would make it look worse. This is particularly a problem if the group is small since the scores of just one student can affect the results for the whole group.

Consider the following example of posttest scores for a Title I class with four students.

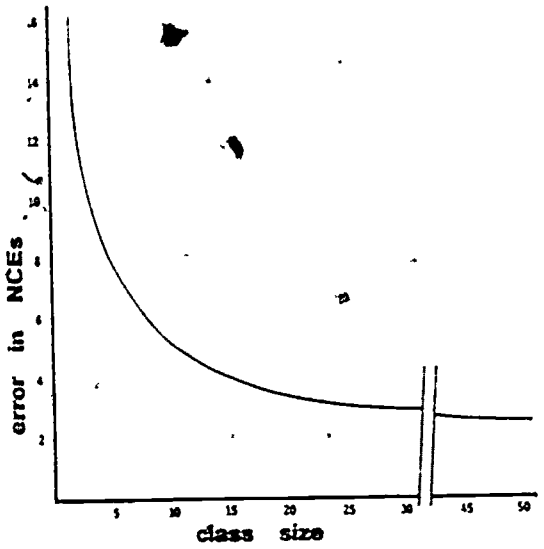
	23	
	37	
	16	
	24	
Total =	100	Average score = 25 NCEs

Now suppose that the first student was ill the morning of the test (first type of error mentioned above) and that the second student never tried to learn during the program (second type of error mentioned above). Both students did less well than they should have on the test. In this case the scores might have been:

	13	
	27	
	16	
	24	
Total =	80	Average score = 20 NCEs

In a class of four students, these two errors had a large effect on the posttest average for the group. In a class of 25 students, the effect would not have been very big.

You can estimate how error affects the accuracy of your results by referring to the following graph (see Technical Note 1):



As you can see, the accuracy of an NCE gain varies depending on the number of students in a Title I class. The amount of error decreases as class size increases. You can also see that the size of the error begins to become more constant when the class size is 15 or more. The error sharply increases as the class size becomes smaller than 15. Note that class size here refers only to students included in computing the gains, that is, students for whom there were both pretest and posttest scores.

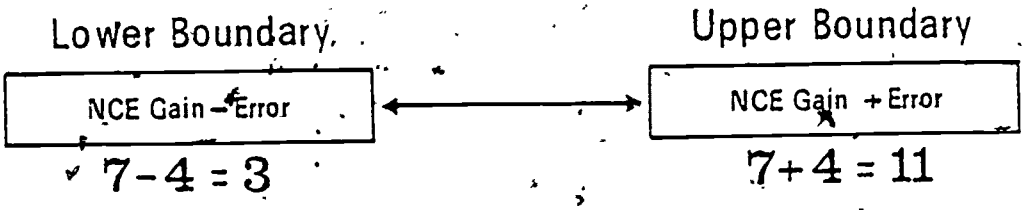
How do you know where your true gain lies?

While you don't know what the true gain is exactly, you can estimate a band within which you can be quite confident the true gain would fall. To construct this band, you can use the table below. The confidence band is your project's gain, "give or take" the amount of error given for your class size. The table is based on the preceding graph.

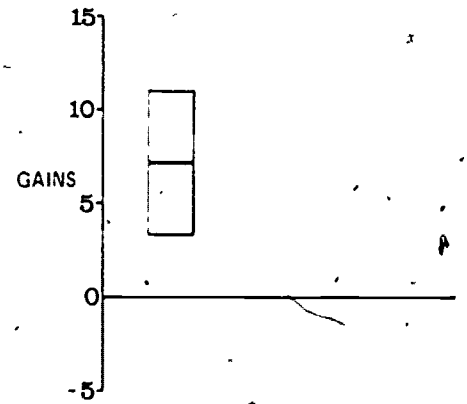
"Give or Take" Table

Class Size (N)	Error (NCEs)	Class Size (N)	Error (NCEs)
2	16.0	21	3.6
3	11.3	22	3.5
4	9.2	23	3.4
5	8.0	24-25	3.3
6	7.2	26	3.2
7	6.6	27-28	3.1
8	6.1	29-30	3.0
9	5.7	31-32	2.9
10	5.3	33-34	2.8
11	5.1	35-37	2.7
12	4.8	38-40	2.6
13	4.6	41-43	2.5
14	4.4	44-47	2.4
15	4.3	48-50	2.3
16	4.1		
17	4.0	75	1.9
18	3.9	100	1.6
19	3.8	150	1.6
20	3.7	200	1.1

Suppose that the gain for your Title I group of students is 7 NCEs and you have scores for 17 students. Using the table, you would find that for a class of 17 students, the error for this gain is about 4 NCEs. In this case, the confidence band is formed by subtracting the error, 4 NCEs, from the NCE gain of 7 to get the lower boundary of the band. The upper boundary of the confidence band is found by adding the error to the gain.



You can be quite confident that the true gain of the program was somewhere between 3 to 11 NCEs. The program did have a positive effect. The confidence band can be pictured graphically as:



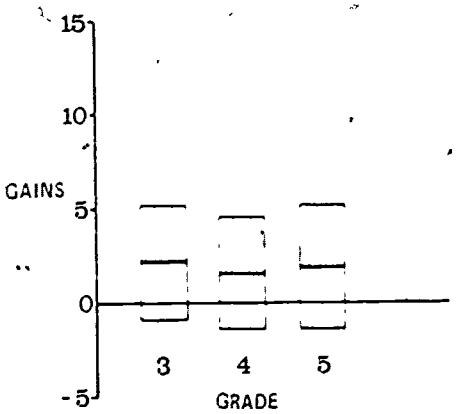
If the observed gain had been 3 NCEs, however, the confidence band would have been -1 to 7 NCEs. In that case, you could not be sure whether the true gain was greater than zero. That is, with a gain of 3 NCEs with an N of 17 you cannot be sure whether or not your program was effective in boosting the achievement of your students any more than would have been expected with just regular schooling.

Example--Confidence Bands for NCE Gains

The Title I coordinator thought that the gains from Wilson's reading program were rather small and wanted to see how accurate the results were. Using the "give or take" table, the coordinator made up the following worksheet:

Grade	N	Gain	Error	G - E	G + E
3	27	2.1	3.1	-1.0	5.2
4	28	1.3	3.1	-1.8	4.4
5	22	1.7	3.5	-1.8	5.2

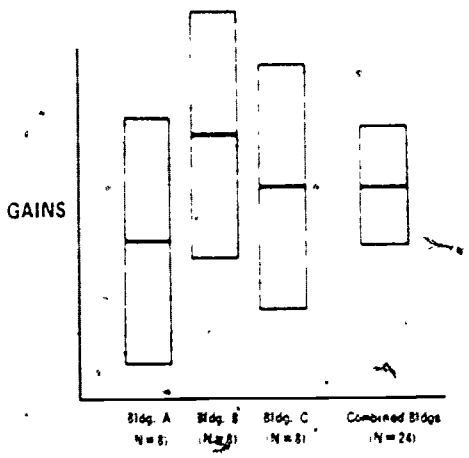
When the confidence bands were graphed, the results looked like:



What should you do if the class size is small?

It should be quite apparent from the "give or take" table that when the class size is small, gains are not very accurate. The confidence band can be quite large with a class size less than 15. One solution is to increase your class size by combining the scores from several classes before interpreting the project's effect. There are several possible ways to combine scores.

Combining Schools. One possibility is to combine the scores from the same grade across schools. If Title I instruction is given to fifth graders in Schools A, B and C, then it would be possible to combine the fifth grade students' scores from these three school buildings. As the figure below shows, this greatly improves accuracy.



Combining grades. If your Title I project is similar across grades, another alternative is to combine the scores from adjacent grades to increase the class size. For example, suppose your Title I project includes the third, fourth, and fifth grades but in the third grade only three students are participating in Title I. Instead of interpreting the NCE gain for only the third grade students, you could combine the students' scores from the third, fourth, and fifth grades. However, the NCE gains for elementary students are generally larger than the gains for students in the junior and senior high schools; so you may not want to combine secondary results with elementary project results.

Combining years. A third possibility is to combine the data across years of the program. If the same teachers, materials, tests, and such have been used in the Title I program for two (or more) years, you can combine the scores for students in each grade over both years. For example, a program with scores for 10 fourth graders in year 1 and 8 fourth graders in year 2 could combine the two sets of scores for a total N of 18, reducing the confidence band by about a third.

When determining the NCE gain for combined scores, do not average the NCE gain from one class with the NCE gain from another class by adding them and then dividing by 2. This kind of averaging will produce an incorrect gain since the original gains were based on different class sizes.

There is a short cut method for combining scores if you have already computed the gain for each class you want to combine. Simply use the following worksheet to multiply each gain by the number of students (N) used to compute it. Divide the sum by the total N to get the overall gain.

Year	School	Grade	N	Gain	N x Gain
			Total: _____		Total: _____

$$\text{Overall gain} = \frac{\text{Total N} \times \text{Gain}}{\text{Total N}} = \_\_\_\_\_$$

Example--Combining NCE Gains for Interpretation

The language arts program in Grades 7 and 8 at Centennial Junior High has only a few students at each grade level. The Title I coordinator decided that the combined results of the two grades would be more meaningful than for either grade separately. The coordinator also decided to lump the data from last year and this year together since the program had been implemented in the same way and the same test was used both years. The results were as follows:

Year	School	Grade	N	Gain	N x Gain
79-80	Jr Hi	7	2	12.1	24.2
79-80	Jr Hi	8	5	6.4	32.0
80-81	Jr Hi	7	4	4.7	18.8
80-81	Jr Hi	8	3	-0.5	- 1.5
			14		73.5

$$\text{Overall gain} = \frac{73.5}{14} = 5.3 \text{ NCEs}$$

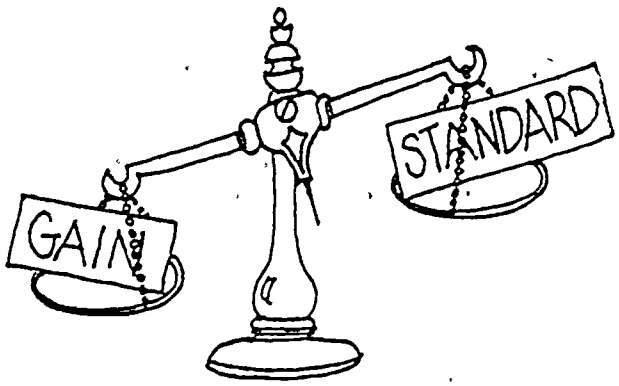
Confidence band = 5.3 $\pm$ 4.4 NCEs

The overall gain for the language arts program (based on 14 students) was much more encouraging than any of the results based on fewer students. Little confidence can be placed, for example, in the 12 NCE gain at grade 7 since the error band for only two students is 16 NCEs.

SECTION TWO

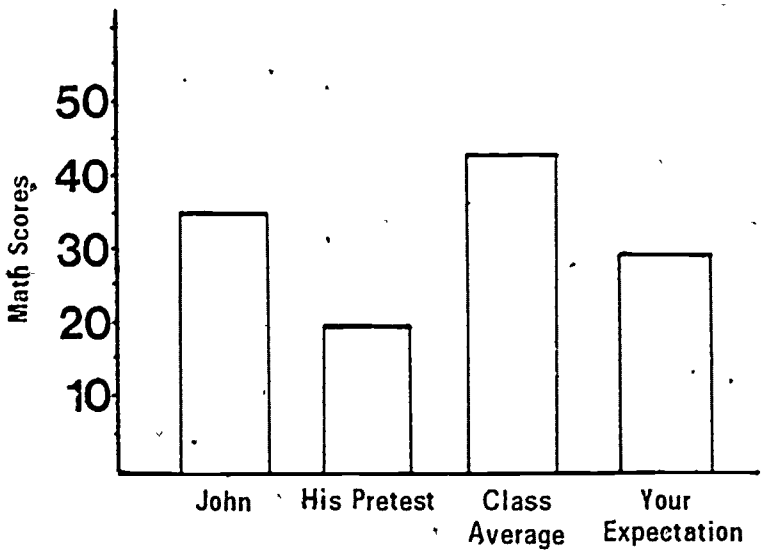
COMPARING PROJECT GAINS TO STANDARDS

The focus of the last section was on how to determine whether your Title I students learned more than they would have without Title I instruction. That is fine, but it does not go far enough. You still do not know how good or how effective the project was.



You have probably encountered a similar problem in trying to attach meaning to a student's test score. Take, for example, a student, John, who scored 35 on a 50-item math test. You do not know whether 35 is a good score or bad score without comparing it to some standard. You would want to know how John did on the test before taking the class, how other students in his class did, how other students at his ability level did, how you think he should have done. Using such information, you can construct standards for judging the worth of a score of 35.

We might find, as in the example below, that while John did not do as well as the class average, he did improve substantially over his performance before instruction and did exceed your expectations for him.



How do you know how good your gains are?

You cannot tell how effective a Title I program is by simply looking at the size of the NCE gain. You might not expect the same size of gains for your secondary program that you would for your elementary program. You might not expect the same gains for a limited, poorly funded project that you would for one that involved intensive remedial instruction. To determine how good or bad the program was, you must have something to compare it to; you must have standards.

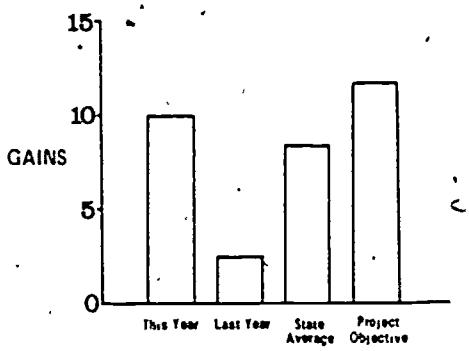
The following table suggests possible sources of standards that you may find helpful and presents the questions based on a third grade project. Each comparison asks something different about that example project.

Question	Possible Standards
"How did my Title I third grade program compare to third grade Title I programs across the state?"	Average gains for Title I projects across state for the same grade and subject
"How did my third graders do compared to last year?"	Performance of same Title I students last year
"How did my third grade program do compared to my third grade program last year?"	Gains from last year's project at each grade
"How did the third grade program do compared to the second and fourth grade Title I classes?"	Project gains at adjacent grade levels
"Did the third grade program make the project goal of five NCEs gain?"	Project objectives based on professional judgment

The process of making judgments about your program is, by nature, subjective. The comparisons that you select will provide rough standards by which to judge your program, but the comparisons cannot be made without qualifications:

- Use several comparisons; do not rely on just one.
- Temper your judgments based on the similarities and differences between the chosen standards and the program being evaluated.

In the example below, the results for a Title I project were compared to three standards: a) the project gain from last year, b) the average gain across the state for that subject and grade, and c) the goal or objective stated in the project application. The gain meets the project objective and is close to the gains observed for the average Title I program. In addition, there appears to be improvement over last year's results.

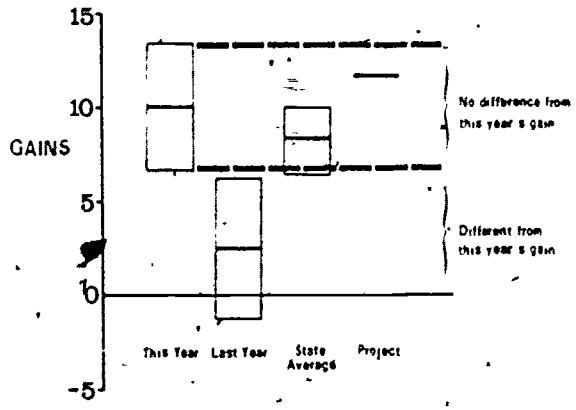


How big a difference is important?

Your project's gain will rarely be exactly equal to the standard. How big must the difference be between the observed gain and the standard before it is worth considering? Is a difference of 2 or 3 NCEs important? Probably not. How about 4 to 5? You must consider the accuracy of both gains before answering that question.

In the first section the point was made that there is always some error involved in determining project gains. Constructing a confidence band about a gain using the table given in Section One provides a way of estimating the effect of error for each gain. If the confidence bands of two gains overlap, you cannot be confident that gains actually differ. Alternatively, if the bands do not overlap you could conclude with some certainty that the gains do differ.

In the example above, for instance, the results for a Title I project were compared to three standards. As shown in the figure below, a confidence band can be constructed around each gain based upon the number of students for each. Since the state average was based on the scores of many students, the band is very small. It did not make sense, however, to construct a band around the project objective since it was based on professional judgment rather than computed from student test scores.



You would conclude from this example that the project was effective in promoting the achievement of educationally disadvantaged students. The project objectives were met and the results are similar to those observed for other Title I projects across the state in that subject and for that grade. There was, however, a substantial improvement over the results from last year. It would be interesting to explore differences in the way the program was implemented each year to try to explain this finding.

It should be apparent that when gains are computed on fewer than 15 students, the confidence bands are quite large and only very large differences between gains can be detected. Project gains based on only a few students can fluctuate greatly. As discussed in SectionOne, the gains for several grades or several buildings can be combined to reduce the size of the confidence bands.

In addition to looking for large differences between project gains and standards, we can watch for consistency in the results. A large gain relative to what was expected seems more reasonable if similar effects were observed at several grade levels or across several years.

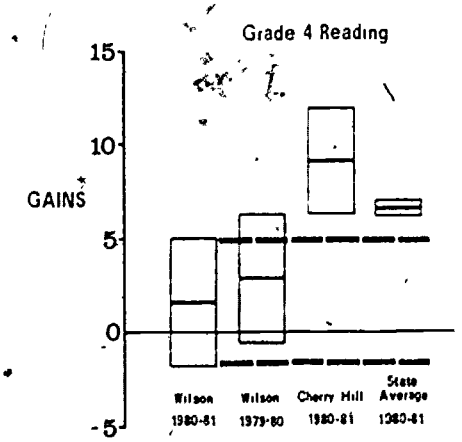
Example--

The Title I coordinator decided that while Wilson's gains in reading seemed discouraging, perhaps it was not unreasonable to expect small gains for this type of program. Several gains were selected for comparison: Wilson's results from last year, Cherry Hill's results from this year, and the state average. To avoid the temptation to make too much of small differences in gains, the coordinator graphed the confidence bands around each gain.

Grade 4 Reading

	N	Gain	Error
Wilson 1980	28	1.3	3.1
Wilson 1979	25	2.6	3.3
Cherry Hill	32	8.4	2.9
State Average	893	6.4	.5

The graph suggests that Wilson's gains at the fourth grade did not differ over the two years of the program and were clearly lower than Cherry Hill's and the state average.

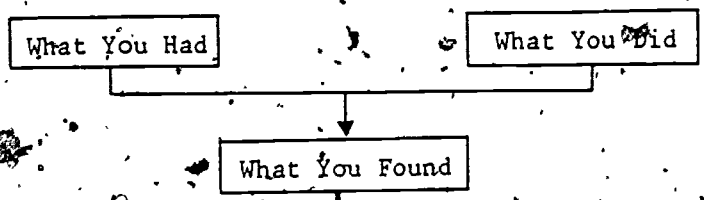


SECTION THREE

INVESTIGATING THE RESULTS

So far you have reviewed what an NCE gain means and how to interpret it by comparisons to other standards. The purpose of this third section is to discuss the next steps that you might want to consider. Before discussing these steps, it is necessary to set the stage for the discussions by describing a very simple method for investigating a Title I project.

The Model. In describing complex activities such as a Title I project, it is often useful to develop a general conceptual scheme that will help focus attention to relevant characteristics of the project. For this section, you will be refining to the very simple model in the following figure:



Very briefly the three components of the model are described as follows:

1. What you had

   This category concentrates on describing what you started with before the project actually began. This includes descriptions of the individuals (students and instructors), environment (school, community, and classroom), and other circumstances such as support or resources available. These factors will have some effect on the impact of your program.

2. What you did

   "Title I instruction" is a very generic term that incorporates a wide variety of materials and techniques. This category concentrates on describing in more detail what actually occurred during the project. This project description will help you focus on the critical features of your project.

3. What you found

   Until now this guide has concentrated on one indication of a project's impact--the change of NCEs from pretest to posttest based on one achievement test score. Although this gives a general picture of a project's impact, you may need to obtain more detailed information about the impact.

   The final element of this model is the connection between "What You Had" and "What You Did." This connection is meant to emphasize the fact that besides each of these individually affecting the results of

a project, the match between the two is also important. For example, a project could be an excellent project for some students but not for others because of differences in their backgrounds.

The outline of the section.

The third section follows an outline of investigation.

Step 1: Should there be a next step?

Investigating the results of a project can be time consuming. Before finding and allotting the resources necessary for the investigation, you should carefully analyze whether it is worth it. Is there interest? Will the information be useful? Is there a potential for changing the project?

Step 2: Substantiating "What You Found"

Jumping into a thorough investigation based on one piece of information, an NCE gain, is questionable. Before embarking on an investigation, you should collect substantiating evidence for the NCE gain.

This discussion offers some alternatives that also reflect the impact of the Title I project and hopefully verify the observed NCE gains.

Step 3: Focusing the investigation. (Defining evaluation questions)

Deciding whether to start with contextual variables ("What You Had") or implementation variables ("What You Did") can be confusing. A few suggestions are offered that might help you decide.

Step 4a: Investigating "What You Had"--(Needs Assessment)
or
Step 4b: Investigating "What You Did"--(Implementation Evaluation)

These two discussions suggest some kinds of information that you might collect to better understand why your Title I project had the impact it did.

Step 1: Should there be a next step?

Now that you understand NCE gains better you must decide whether some actions should be taken. In Title I evaluations the information can lead to three possible situations:

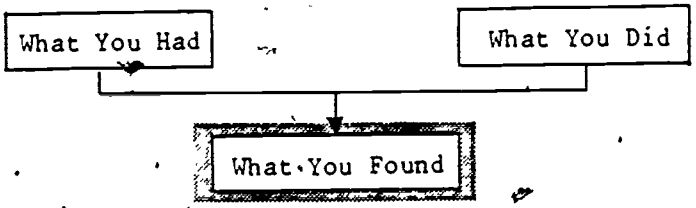
1. You could find that the Title I project produces very small or even negative NCE gains. NCE gains could be satisfactory but not up to expectations. You may now want to modify the existing program to achieve higher gains or decide if you should abandon the approach and adapt or adopt another Title I instructional approach.

2. The NCE gains from the Title I project can be satisfactory to everyone involved. Resources and energies can be put into continuing the program as it is. Although a thorough investigation may be unnecessary, you will still probably be collecting some information for program management and implementation reasons.

3. You may find that the NCE gains are very high compared to what was expected or that achieved by other projects. In this case you may be interested in identifying the particular strengths of the project either for modifying other projects in the district or for disseminating the project to other schools.

The first and third situations, although stimulated by opposite extremes of results, demand that some action be taken. In the first situation, you are concerned that the project get up to par; in the second, as an educator you are concerned with sharing what works well. The second situation requires no further action. It would be naive to avoid understanding why a project works, however. It would be wise to collect and document what information exists in case it needs to be used in the future. Therefore, the answer to the question of whether there should be a next step is usually yes. The three situations differ in how much information is needed and how soon.

Step 2: Substantiating "What You Found"



"Look before you leap" is a good motto for a Title I evaluator. An NCE gain is only one piece of information from one achievement test or even subtest. Now you should be able to see why it is so important that the test matches the project's curriculum, is sensitive to the project's effect, and is administered properly and according to the evaluation model's guidelines. Violations of these assumptions could result in extremely negative or positive results or false satisfaction from an "acceptable" NCE gain.

Before you decide to proselytize a project or drastically alter it, verify the evaluation results. Although this certainly means checking to see that everything was done properly in the evaluation, you should also look for substantiating evidence of the project's results. If other information substantiates the NCE gain then you can be fairly confident about your decision to continue to investigate causes. If, on the other hand, the different sources of information produce conflicting information, then you must investigate these conflicts first.

1. Other Standardized Tests

Many Title I students take other tests outside the Title I project. For instance, the school district may administer a nationally standardized or locally developed test to all students. Although the test for the Title I students is chosen because it most closely matches the instruction, the results from other testing should not conflict drastically.

2. Teacher-Made Tests and Student Skills Checklists

During the project, the Title I teacher probably administers small formal or informal tests to keep a record of the student's progress. Comparing the record of the achievement on these tests will aid in substantiating the NCE gains.

3. Standardized Curriculum Tests

If a project has adopted a curriculum from either a large publishing firm or another project that is disseminating its materials, progress tests are often a part of the curricular materials. Often manuals are provided on how to interpret results from these tests. These can be used in the same way that teacher-made tests are used to verify an NCE gain.

4. Regular Classroom Test Results

The Title I evaluator should also be aware of the achievement results of the Title I student in the regular classroom. These tests may be more difficult than those given in the Title I instruction and often are given in more stressful situations than the Title I tests but they can be very useful as substantiating evidence.

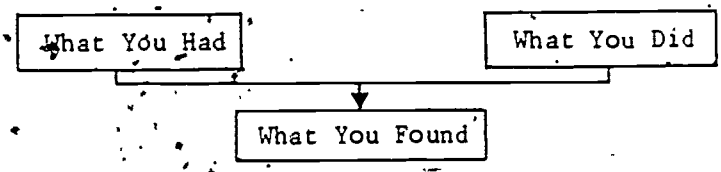
5. Gut Feelings

Title I evaluators often concentrate on collecting objective information which is typically in the form of test results. Often there is much to be gained by documenting what the teachers and students feel about the impact of the Title I project. This information may even identify some additional benefits (or detriments) of the Title I project that were not originally thought of to be a part of the program. Care should be taken in trying to provide some structure to the format of collecting subjective information. That is, just a "What do you think of the Title I project?" will elicit a wide range of responses. Some could be useful; some not. If there are specific areas you are interested in, then address these in your collection of responses.

To expect one hundred percent agreement from all sources is unrealistic. Conflicting results can be caused by some sources not being as exact as others or by one source keying in on a factor that other sources ignore. These conflicts in themselves can offer information about how a project functions. You will probably feel more certain in your investigations if you start out with most of the evidence substantiating either negative or positive results.

Step 3: Focusing the investigation--(Defining evaluation questions)

After you have decided you want to investigate the reasons behind the gains the project showed, you must decide where to begin. Earlier we referred to a model:



If you're limited in resources, you must decide where to concentrate. There are a couple reasons why you would concentrate on one area before the other.

1. How well do you know the territory?--(Needs assessment)

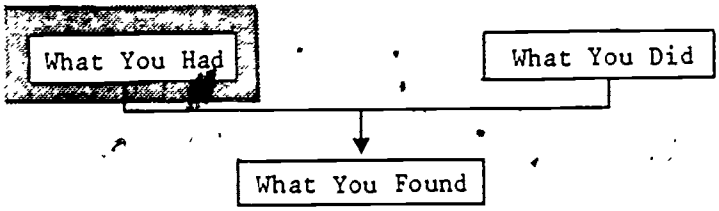
The "What You Had" component in the model includes most of the contextual variables behind your project--what kinds of students are served, or what training have the instructors had, for instance. An investigation of this are may show that the problem is not that the project is poor, but that it doesn't match the students or the situation in which it was used. In evaluation the process of describing a situation and assessing the strengths upon which to build and the weaknesses to correct is called a "needs assessment." A thorough needs assessment is essential to the success of any project. Investigations in the area "What You Had" do not usually identify areas for changes in the project, but they can identify variables that affect the success.

2. Are you concerned with adjusting the project?--(Program implementation)

If you have a thorough understanding of "What You Had" then you should investigate the area called "What You Did." Investigations here are referred to as "implementation evaluations" and study how the program actually ran (as opposed to how it was designed to run). The concern here is what refinements to the project can be made to improve the results. For example, you may be interested in how much actual instruction time each child received or if the Title I instruction was coordinated with the instruction received in the regular classroom. These kinds of variables are most apt to be changed from year to year as a Title I project is refined.

The final two parts of this section suggest some general questions to answer in getting to understand your project better. They are by no means meant to be exhaustive but are to help you in deciding what areas of your program you might start to investigate. Once you have decided, you should refer to other materials that more completely describe needs assessments or implementation evaluations.

Step 4a: Investigating "What You Had"--(Needs Assessment)



1. What was the average pretest achievement?

Sometimes by concentrating on NCE gains you may forget to look at pretest scores as an indication of where the students began. Although Title I students are below average in achievement, you would certainly agree that there are different levels of "below average." A project that works well for students at the 25th percentile may not be as successful for students who average at the 10th percentile. This is important to remember especially if you're considering adopting a program that has worked well for another district. You should consider whether there is any difference between the students' achievement levels and if that would affect the effectiveness of the project.

2. How proficient are the students in English?

All the students entering a Title I program may not have the same degree of ability to deal with basic English comprehension. Achievement in reading or matnematics is often adversely affected by language problems.

3. What kind of student "turnover" is there?

The success of a Title I project can be adversely affected if the Title I student population has a large percentage of students who are moving in and out of the school during the year. Some initial indication of this problem can be obtained from the evaluation report by comparing "Participation," the number of students who received any Title I instruction and "N," the number of students with both pretest and posttest scores. A high "Participation" with a low "N" may be an indication of a program that has a high turnover in students.

4. Do the Title I students have a high absentee rate?

If students are only in school say three of five days a week, then the amount of instruction they receive is substantially reduced. This could be reflected in the evaluation of the project's impact.

5. Are the students motivated?

Title I evaluation concentrates on achievement impact; however, whether a student does well is affected by the students' motivations and self-concepts.

6. How are students chosen to be in Title I?

If you are comparing your program to another, you should study how the students are chosen for the projects. If selection procedures vary, the projects may be designed for different types of students.

7. What kinds of demands does the project require of Title I instructors?

These demands may range from special training in remedial instruction to the amount of time required by the project to be spent with each student.

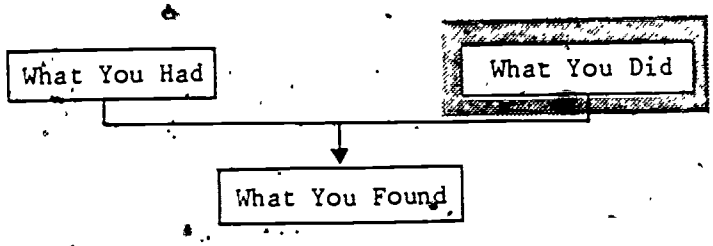
8. How would you describe the school environment?

Schools differ on the types of students served, instructional staff expertise, or administrative and classroom teacher support for the project.

9. How would you describe the community that the school serves?

Besides the school, you may also need to describe the community and the parents to identify ways in which they indirectly or directly influence the Title I project.

Step 4b: Investigating "What You Did"?--(Implementation Evaluation)



1. What general information is available about the project's implementation?

From your records you can get some very general information about how the program was implemented. This includes: project setting, hours per week of instruction for the typical student, the project length in weeks, and the student-to-instructor ratio. These may be different from how the program was originally designed.

2. What specifics about the project's curriculum are important?

- How does the Title I program complement the regular classroom work? Is there a coordination between the two?
- What specific materials are used in the program?
- What are the project's goals and objectives?
  - How many are there?
  - How many general areas are covered?
  - At what level are they written? (How specific are they?)
- How were the instructors trained?
  - Topics of inservice training
  - Number of days of inservice training
- What staff levels and FTE levels are required?

3. How was the project actually implemented?

- Are there specified structured teacher behaviors for the project (positive reinforcements, specific questioning techniques, etc.)?
- Are there distinctive program features such as volunteers, peer tutoring, home visitation, parental involvement?
- Where does the instruction occur (in a special room, in the classroom)?
- Do the procedures complement the objectives of the project?
- How do teachers work with the students (individually, large group, small group, combination)?

- How is student progress monitored?
- When does instruction occur (morning, afternoon, near recess, during a particular course)?

4. How much actual instruction does a Title I student receive?

- How much student time is spent in Title I per day and week?
- How many weeks is the student usually in Title I?
- What are the criteria for exit from the program?
- What are the students missing in regular instruction?
- How much time is spent in testing, getting to and from the Title I classroom, in administrative detils, etc.?

EPILOGUE

Is that all there is to interpreting evaluation results? The answer is, of course, "No." This guide has taken you through some suggestions on how to make NCE gains more meaningful and how to interpret the gains you get; but you can certainly expand the effects of a Title I program beyond achievement test results. For instance, there is the effect of the program on the students' attitudes to school and about themselves and others. Besides the students, the program may also affect others such as non-Title I students, teachers, administrators, or parents.

This guide will, however, stop here and acknowledge the existence of interpretation beyond this level. This in no way should be mistaken as a sign that these effects are not important. In fact, they are the very essence of a comprehensive evaluation plan.

TECHNICAL FOOTNOTES

1. The effect of sampling error or gains can be estimated using the standard error of a difference. The standard error is the expected standard deviation of gains if a large number of samples were drawn from the national population of Title I students. The values in the graph represent the standard error of a difference (gain) for varying class sizes. The formula used was:

$$SE_{diff} = \sqrt{\frac{s_x^2 + s_y^2 - 2r_{xy}s_x s_y}{N - 1}}$$

where

SE_{diff} = standard error of a gain/difference

s_x^2 = variance of pretest scores

s_y^2 = variance of posttest scores

r_{xy} = correlation between pre- and posttest scores

N = number of students

(Horst, Tallmadge and Wood)

Since standard deviations and correlations are generally not available for Title I evaluations, estimates of 16 NCEs and .50 respectively, were made based on the data for a number of Title I projects.

2. The confidence band corresponds to a 68 percent confidence interval since the tabled values represent one standard error. This level of confidence seemed appropriate since it increased the chance of identifying small gains that would have been missed with a 95 percent confidence band constructed from about two standard errors. This procedure is comparable to that recommended by test publishers for interpreting a student's score or a group average. The publisher provides the standard error of measurement for the test and the teacher constructs a confidence band around the score or group average.

3. Technically, the standard error of a difference between gains should be used to estimate the effects of error in comparing gains. The formula is:

$$SE_{diff} = \sqrt{SE_1^2 + SE_2^2}$$

This would be complicated and difficult to estimate. It can be shown, however, that the sum of the two standard errors (SE_1+SE_2) is greater than one standard error of a difference (SE diff) and less than two. Thus, combining the standard errors from the two gains results in a band that is between the 68 percent and 95 percent confidence intervals and seems to provide a good approximation to the proper statistic.

Graphically, one simply constructs the two confidence bands for the gains separately. If the bands overlap, the difference between the gains is not considered significant. This is essentially the same procedure recommended by test publishers for interpreting student or group profiles, but using the standard errors for gains rather than the standard error of measurement.